

AD-A203 113

Literature Survey Part I: Assessment Technology

Patricia A. Harris and Evelyn Hendrix

Allen Corporation of America

for

Executive Development Research Group
Thomas O. Jacobs, Chief

Manpower and Personnel Research Laboratory
Newell K. Eaton, Director

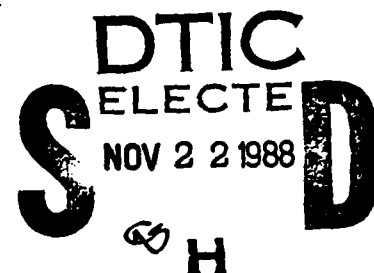


U. S. Army

Research Institute for the Behavioral and Social Sciences

September 1988

Approved for the public release; distribution unlimited.



88 11 21 093

U.S. ARMY RESEARCH INSTITUTE

FOR THE BEHAVIORAL AND SOCIAL SCIENCES

**A Field Operating Agency Under the Jurisdiction
of the Deputy Chief of Staff for Personnel**

EDGAR M. JOHNSON
Technical Director

JON W. BLADES
COL, IN
Commanding

Research accomplished under contract
for the Department of the Army

Allen Corporation of America

Technical review by

Donna C. Angle
Paul van Rijn

This report, as submitted by the contractor, has been cleared for release to Defense Technical Information Center (DTIC) to comply with regulatory requirements. It has been given no primary distribution other than to DTIC and will be available only through DTIC or other reference services such as the National Technical Information Service (NTIS). The views, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other official documentation.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified			1b. RESTRICTIVE MARKINGS -	
2a. SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.	
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE -				
4. PERFORMING ORGANIZATION REPORT NUMBER(S) ARD-TR-86-3			5. MONITORING ORGANIZATION REPORT NUMBER(S) ARI Research Note 88-90	
6a. NAME OF PERFORMING ORGANIZATION Allen Corporation of America	6b. OFFICE SYMBOL (If applicable) ARD	7a. NAME OF MONITORING ORGANIZATION U.S. Army Research Institute for the Behavioral and Social Sciences		
6c. ADDRESS (City, State, and ZIP Code) 209 Madison Street Alexandria, VA 22314		7b. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333-5600		
8a. NAME OF FUNDING / SPONSORING ORGANIZATION see 7a.	8b. OFFICE SYMBOL (If applicable) PERI-RO	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER OPM 41-83		
8c. ADDRESS (City, State, and ZIP Code) see 7b.		10. SOURCE OF FUNDING NUMBERS		
		PROGRAM ELEMENT NO. 6.37.31	PROJECT NO. A792	TASK NO. 4.5.5
		WORK UNIT ACCESSION NO. 4.5.5.C2		
11. TITLE (Include Security Classification) Literature Survey Part I: Assessment Technology				
12. PERSONAL AUTHOR(S) Patricia A. Harris and Evelyn K. Hendrix				
13a. TYPE OF REPORT Final Report	13b. TIME COVERED FROM 3-85 TO 6-88	14. DATE OF REPORT (Year, Month, Day) 1988, September	15. PAGE COUNT 89	
16. SUPPLEMENTARY NOTATION Thomas O. Jacobs, contracting officer's representative. Literature survey conducted with the assistance of John Alderson and Mary Ann Alderson. (OVER)				
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)	
FIELD	GROUP	SUB-GROUP	Cost/Benefit Analysis Assessment Procedures	
			Executive Development Assessment Technology	
			Validity Measures	
19. ABSTRACT (Continue on reverse if necessary and identify by block number) A literature survey of assessment technology and application was conducted to identify techniques that might be useful to an executive development program for future leaders of the U.S. Army. Literature was sorted into five categories: assessment purpose, procedures, validity, cost/benefit analysis, and assessor qualifications. Abstracts of 27 validity studies were prepared and presented in a summary table. A second summary table presented information on level, validity, purpose, costs, and assessor level for 16 assessment programs. The majority of the literature was found to be based on mid-level assessment for industrial requirements. Research and application is currently moving toward situational testing rather than psychometrics. Assessment center techniques continue to be popular in medium and large firms. Longitudinal studies of predictive validity for assessment center techniques show acceptable validity measures. Cost/benefit analyses of assessment techniques are hampered by a lack of individual benefit metric. <i>Keywords: Cost Analysis, Executive Development, Office Personnel, Research, Survey.</i>				
20. DISTRIBUTION / AVAILABILITY OF ABSTRACT <input checked="" type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL Thomas O. Jacobs			22b. TELEPHONE (Include Area Code) 202/274-9045	22c. OFFICE SYMBOL PERI-RO

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

ARI RESEARCH NOTE 88-90

16. Supplementary Notation (continued)

Monitoring also by Office of Personnel Management, Workforce Effectiveness and Development, Office of Training and Development, P.O. Box 7230, Washington, D.C. 20044.



Accession For	
NTIS GRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

FOREWORD

The U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) is actively pursuing research into the long-term development of the Army's future executive leaders. One element of the research is focused on assessment and development technologies that may have utility for the Army's executive development program.

This report presents a survey of the literature in the area of assessment and includes both academic research and technologies in current use in industry. Information is presented on assessment purposes, procedures, validity, cost/benefits, and assessor qualifications -- topics of on-going interest to the research effort. A companion volume will present the results of a similar survey of theories and technologies related to executive development.

LITERATURE SURVEY

PART I: ASSESSMENT TECHNOLOGY

EXECUTIVE SUMMARY

Requirement:

To conduct a survey of assessment literature, particularly regarding current assessment methodologies in use in industry that might be appropriate at the 10- and 20-year mid-course corrections in Army career paths.

Procedure:

Literature sources were identified by key-word searches of computerized data bases. Selected cites were reviewed and sorted into five categories: assessment purpose, procedures, validity, cost/benefit analysis, and assessor qualifications. Abstracts of 27 validity studies were prepared and presented in a summary table. A second summary table presented information on level, validity, purpose, costs, and assessor level for 16 assessment programs. One on-going assessment program was described in depth.

Findings:

The majority of the literature surveyed was found to be based on industrial requirements, and did not reflect differences at the 10- and 20-year career points that are of particular interest to the Army. Assessment centers continue to be a popular method for mid- and large-sized companies. Recent longitudinal studies have shown acceptable validity levels for performance prediction based on assessment center techniques. The general trend in assessment is toward situational testing, rather than psychometrics. Cost/benefit analyses of assessment processes have not yet been formalized, primarily because metrics for individual versus organizational benefits have not been identified.

Utilization of Findings:

The survey confirmed that assessment techniques with acceptable validity measures are being used in industry. The utility of these techniques for assessing future Army leaders needs further investigation.

LITERATURE SURVEY PART I: ASSESSMENT TECHNOLOGY

CONTENTS

	Page
INTRODUCTION	1
Organization of the report	1
THEORETICAL BASES TO ASSESSMENT	2
Trait Theory versus Situational Theory.	2
Interactive Approaches	2
PURPOSES OF ASSESSMENTS	3
ASSESSMENT PROCEDURES	4
Nomination Process	4
Psychometric Tests.	4
Feedback Techniques	8
VALIDITIES OF VARIOUS ASSESSMENT PROCEDURES	8
ASSESSOR LEVEL AND QUALIFICATIONS	10
Assessor Level	10
Assessor Qualifications	13
RELATIVE/COST BENEFITS	14
Cost Variables	14
Start-up Costs	15
Benefits	16
Quantitative Formula	18
SUMMARY OF LITERATURE SURVEY FINDINGS	19
REFERENCES	24

CONTENTS (cont.)

Page

APPENDIX

A: Validity Studies Associated with Assessment in General	29
Index to Appendix A	29
Validity Studies Associated with Assessment in General	30
B: Validity Studies of Specific Assessment Exercises	55
Index to Appendix B	55
Validity Studies of Specific Assessment Exercises	56
C: In-Depth Look at one Management Assessment Program	74

INTRODUCTION

The U.S. Army is vitally interested in both the early identification of executive potential and the development of that potential at a more rapid and predictable rate. These are the basic components of an executive development program being designed to respond to the increasingly complex leadership skills required of General Officers and Senior Executive Service members. A systematic identification process is seen to include formal assessment techniques that could be incorporated into established mid-course corrections at 10- and 20-year career points. Formulating a development strategy will involve identifying the dimensions of executive ability responsive to developmental guidance and the sequence of experiences most likely to result in growth toward known executive capability requirements.

The executive assessment and development needs facing the Army are similar to those confronting other complex organizations. The academic and corporate communities have been addressing those needs for well over two decades through theoretical and applied research, assessment centers, and developmental packages. The result is a large body of knowledge and experience that could be useful to the Army. The purpose of this research was to review the literature in both assessment and development, and to seek findings that might be relevant to a course of action for the Army's senior leadership. The literature survey was supplemented by inputs from practitioners and researchers whose expertise would add depth, breadth, and evaluation to the information documented in the formal literature.

The literature survey is reported in two parts. This volume is a review of assessment technology. Part II, the Executive Development Report, will be presented in a separate volume.

Organization of the Report

The results of the survey of assessment literature were sorted and integrated into five categories:

1. purposes of assessments
2. assessment procedures
3. validity of the assessment procedures including predictive validity at different career points in time
4. assessor qualifications
5. relative costs versus benefits of various assessment procedures at executive levels.

Each topic is presented as a separate section, following a brief background discussion on the theoretical bases for assessment. Abstracts and validity information for 27 selected studies reported in the literature are included in an appendix.

THEORETICAL BASES TO ASSESSMENT

Trait Theory versus Situation Theory

Bowers (1973) provided a name for the two major theoretical threads in assessment research: trait versus situation. Both have strong proponents, and both are important to current assessment technology. A trait is defined as any enduring way in which one individual differs from another (Guilford, 1959). Traits are viewed as predispositions to respond in a certain way in many different kinds of situations. Thus, trait theorists attempt to explain the consistencies of human behavior.

An assessment approach based on trait theory would search for the underlying traits related to effective performance. Once the basic traits that represent sources of desired behaviors have been identified and quantified, the assessment process is structured to select individuals having the greatest potential for producing the desired behaviors. Simply put, find the factors you want to assess for the best leaders so that you can select those having the highest potential for leadership.

Mischel (1968) disputed the claims of trait theorists by demonstrating low correlations between some traits and the behavior they purported to measure. His examples included correlations between personality trait measures and behaviors that were rarely found to be above .30. According to Mischel, such measures accounted for only 9% of the total variance in the behaviors, and thus had little practical value in selecting individuals. Rather than traits, Mischel supported social learning theory and suggested that people behave the way they do because they have been reinforced for certain responses in certain situations. He concluded that there was no reason to expect people to have general traits supporting a kind of behavior or to be classifiable along a trait dimension.

Other researchers, including noted authorities such as Dunnette, have continued to support a trait theory approach. In his chapter in the Handbook of Industrial and Organizational Psychology (1983), Dunnette objected to situationally-based predictions. He contended that human traits exist to a consistent and sufficient degree to warrant the assumption that work performance can be predicted apart from the situation in which it occurs. In his view, trait theory is the most valid and efficient basis for prediction.

Blake and Mouton (1982), writing on developing a science of leadership, discussed the implications of the two approaches. If leadership is seen as largely contingent on the situation, a successful leader changes behavior accordingly. If there is a single, most effective style of leadership, a leader must change the environment to make it consistent with principles of behavior. Blake and Mouton appear to have found more evidence to support a single style as the best approach to sound leadership behavior.

Interactive Approaches

A growing number of researchers are proposing a combination of approaches. Bowers (1973), in a critique of situationism in psychology,

discussed the then-current tendency to account for human behavior in situationistic terms. He noted that the effect of this trend was to substitute a behavioristic explanation of personality for an excessively trait-based approach. Bowers argued that, although behavior is more situation-specific than trait theory acknowledges, situations are more person-specific than is commonly recognized. He offered an interactive account of personality as an alternative.

Osborn et al. (1981) also supported a dual approach to explaining performance variance. They reviewed both individual performance and environmental volatility as factors in executive succession and found that either could be supported, depending on the methodology used. Successions of chief executive officers could be linked to either individual factors such as financial performance, or environmental factors such as corporate divestiture and merger. They pointed to the need for further research on the influence of performance and environmental variables and on the predictive value of specific factors associated with succession.

From this brief review, it appears that variables related to executive performance may exist within the individual, within the situation, or within the interaction of the two. Various combinations of traits and situations will probably account for more variance than either trait or situation variables alone. Thus, an assessment strategy incorporating an interactive relationship between individual and situation would seem to hold the most promise.

PURPOSES OF ASSESSMENTS

Assessments in industrial settings are usually part of a selection process -- either for initial entry or for promotion to a particular position (or level of positions). Assessments are most often conducted in an effort to match an individual's psychological profile with a demand profile for a particular job description. In general, assessment appears to be most popular and valid when the positions for which individuals are being considered are quite different from their current positions. By simulating and testing the challenges of the level of management for which the individual is being considered, management can determine the probability of success in the new position.

In addition to selecting individuals to fill certain positions, the literature contains reference to a broad variety of other uses:

1. predicting job success and promotability
2. determining executive potential
3. evaluating skills and abilities related to current job performance
4. remedial development
5. individual development and personal feedback
6. comparing abilities across levels of managers or across functional groups
7. determining developmental needs related to future positions

About half of the formal assessment centers in the United States are aimed at identifying supervisory potential. The trends in assessment center evaluation are toward earlier identification of potential and assessment for higher levels of management.

Only a few references listed individual development as the purpose of assessment. One leading assessment center recommends to its assessors only that they include a separate paragraph on "development needs" if the assessee has limitations on growth potential that should be highlighted and are correctable. The general lack of developmental assessment probably reflects three factors: (1) identifying individual deficiencies is less demanding than creating a prescription to address them, (2) providing training or developmental experiences that are beyond skills of the evaluator will require involving others, and (3) management philosophies often stress hiring successful executives, rather than growing them.

ASSESSMENT PROCEDURES

Assessment is identifying the raw materials for success that already exist or can be developed in an individual. Traditional assessment has involved three factors:

- nomination processes
- psychometric testing of the individual
- individual and organizational feedback

Nomination Processes

Nomination may be by supervisor, peers, self, or subordinates. If supervisory nominations alone are used to identify assessment candidates, high-potential employees may be excluded either through oversight or design. In addition, attitudes of managers not selected may be affected, particularly if the assessment/development program is a benchmark for promotion. These potential problems are being avoided by either self-nomination or by assessing everyone at a particular level. Over half of the first-level supervisory assessment center positions are now filled through self-nomination, a method strongly recommended by the Equal Employment Opportunity Commission.

Psychometric Tests

Psychometric tests used in assessment are classified as objective, subjective, situational, and projective. Objective tests are psychometric evaluations; subjective tests include perceptions of others gained through observation. Relatively recent techniques include situational exercises, in-basket problems, peer ratings, oral presentations, leaderless group discussions, and group gaming exercises.

While many different types of exercises are still being used, there has been a general decrease in group exercises and a dramatic increase in interview simulation. In this situation, an assessee is placed in the role of a supervisor or manager who must conduct a performance-improvement interview. An assessor or other trained person assumes the role of the subordinate. There is a strong trend toward eliminating the background interview from the assessment process and making it a separate part of a formally-defined promotional system which may also include performance appraisal data.

It is also standard practice not to accept any evidence unless there are two independent sources to identify a behavior or characteristic. For example, evidence must be present in an objective test as well as in a subjectively oriented test or observed directly in the person's behavior. Evidence that is found in three or more sources is generally considered likely to be reliable and valid.

In the following section, the test battery for a current assessment program is presented as an example of the types of tools that are in use. This particular program includes both objective and subjective test categories.

Objective Cognitive Tests

1. Guilford-Zimmerman Verbal Comprehension. A percentile score is indicative of the ability to read, hear, and understand complex materials on first presentation; can also represent ability to speak to or to defend a subject with facility when faced with a diverse or adversarial audience. Score is likely to indicate a more broadly read individual than one who restricts reading to his/her own professional area and light reading.
2. Guilford-Zimmerman Quantitative General Reasoning. Scores are considered indicative of native competence with quantitative materials or continued practice with such materials. Generally, scores are indicative of the ease of grasping complex, abstract formulas, models, and conceptualizations on first presentation.
3. Watson-Glaser Critical Thinking. Scores are indicative of a natural or practiced facility to address problems through logical analysis or constructs, rather than from associations or past experience. The five subscores can be indicative of the relative strengths or weaknesses within logical capabilities. Inference and deduction are classic logical constructs; recognition of assumptions, interpretation, and evaluation of arguments deal with meaningful associations and evaluations within a logical structure.
4. Productive Thinking Test. Scored on three facets of productive thinking: 1) absolute number of quality ideas, 2) number of different aspects of the problem which were addressed (breadth), and 3) depth of insight into which the individual delved. Results can be indicative of overall productive thinking, creative thinking, ability to brainstorm ideas; and, to some extent, flexibility of thinking.

These four cognitive tests evaluate a candidate's intellectual effectiveness. If the scores for an individual are clustered together, it may be inferred that these resources can be drawn on about equally. To the extent that the different test scores are spread out, it may be inferred that the individual will tend to utilize the stronger skills in more situations and for more purposes.

Subjective Tests

1. Guilford-Zimmerman Temperament Survey. This standard personality inventory is useful in determining the relative strengths and direction of several characteristics. The characteristics, when pronounced, are most apt to become a discernible aspect of the individual's workstyle or relationships with others, or they may emerge when the individual is under considerable stress. Interpretation provides descriptive information for the summary categories of work approach and relationships with others, using nine subscales:

General Activity. Location on the high or low end usually indicates a preference for fast-paced, energy-charged work effort or a low-keyed, easy-going work approach.

Restraint. Scores at the high end would be seen as showing serious-minded, deliberate, effort and self control. Low end indicates carefree, happy-go-lucky, impulsive, and excitement loving behavior.

Ascendancy. Measures visible signs of leadership. Those scoring in the high range seen as seeking leadership positions; those at the low end generally following rather than leading.

Sociability. Tendencies toward introversion or extroversion are indicated. High-end individuals likely seen as having many friends and acquaintances and seeking the limelight. Low-end would indicate having few friends and being shy.

Emotional Stability. High equates with evenness or constancy of moods and interests, an optimistic outlook, with feelings of composure and good health. Low is interpreted as having fluctuating moods, showing pessimism, or feeling in ill health, guilty, lonely or worried.

Objectivity. High-end individuals seen as being thick-skinned in the face of criticism. At low end, seen as exhibiting hypersensitivity to criticism, self-centeredness, suspiciousness, and getting into trouble.

Friendliness. High end seen as tolerant of hostility, accepting of domination, and showing respect for others. High friendliness may be indicative of passivity or of taking life in stride. Low end apt to be seen as belligerent, resistant to domination and contemptuous of others.

Thoughtfulness. High end described as reflective, philosophically inclined, observant of themselves and others, and having mental poise. Low end seen as being more interested in overt activity, daily operational activities, and as being more apt to show mental disconcertedness.

Personal Relations. High end seen as tolerant of people and showing faith in social or society's institutions. Low-end individuals seen as hypercritical of people and institutions, suspicious, and indulging in self-pity.

It is standard practice in this assessment program to consider any information from the temperament survey that falls between the 20th and 80th percentile to represent variations of average behavior which may, on occasion, lean in one direction or another. Above the 80th and below the 20th percentiles, the behavior is considered to have a predominating quality.

2. Sentence Completion Test. Assessors look for redundant themes and for statements supporting (or contradicting) temperament characteristics related to work approach and relationships with others. There is no standard for scoring. These tests are used to validate or challenge the objective test information on the same general dimensions.
3. Letter Writing Exercise. This is close to a job sample, with the assessee responding to a hypothetical situation that involves sensitive human relations issues. There are 20 criteria on which the results are judged: legibility, spelling, grammar, clarity of meaning, organization, completeness, length, understanding of relevant issues, strategy employed, avoiding lifting phrases, strength of persuasion, positive or negative attitude, use of power, conciliation, use of praise, use of depersonalized criticism, tactfulness, firmness, overall tenor of letter, overall effectiveness of letter. This exercise seems to give a good indication of how personal relations are handled, as well as the use of strategy and power to achieve goals.
4. Thematic Apperception Test. Interpretations of card-based stories draw on 15 basic human needs identified by Murray (achievement*, endurance*, affiliation*, nurturance*, autonomy*, deference*, succorance*, change*, order, aggression, exhibitionism, heterosexuality, dominance, abasement, and intraception). The test taps the eight marked by * most frequently. Needs are used to identify contributions to work motivation and relationships with others.

After all psychometric tests are completed, scores are assimilated and results are compiled. The assessee learns of the results during feedback interviews.

Feedback Techniques

Feedback techniques at assessment centers include the following:

- professionally led critiques of performance in group activities
- videotaped feedback
- peer group evaluation of in-basket exercises (to share decisions and actions, evaluate reasoning, and broaden repertory of responses)
- feedback interviews

Properly structured feedback sessions can be an important benefit to individuals. The amount and detail of feedback vary greatly, in a pattern that appears to be related to organizational level. Higher-level participants get much more information than lower level personnel. Career counseling and planning discussions are often combined with assessor feedback for higher level participants.

The feedback process is particularly important in the cases of employees who do poorly. Organizations that have made a deliberate effort, through expert handling of the feedback process and alternative methods of advancement (e.g. technical ladders), experience fewer problems.

VALIDITIES OF VARIOUS ASSESSMENT PROCEDURES

One of the major topics explored in the literature survey was the validity of various types of assessments. The specific types of assessments included: psychometric, individual interview, job sample, simulation, and small group. Part of the problem in validating assessment tests as separate entities is explained by Finkle (1983). He quotes Dr. E. K. Taylor's reference to different assessment center approaches as "variations on a theme," and points out that the use of several assessment techniques is a common practice. The validity studies in the assessment literature reflect this pattern. It should also be noted that in some instances actual statistical validities are given, i.e., correlation of the assessment measurements with a criterion. However, other references include only general comments, usually opinion-based, regarding the validities of particular approaches.

At least one attempt has been made to evaluate the predictive validity of independent selection procedures (Korman, 1968). Many of the studies reviewed used only one psychometric type, e.g., cognitive ability or personality inventory, and the validities generally were found to be quite low. In the few studies where psychometric tests were part of a larger battery, Korman found that judgmental ratings were better predictors than psychometric tests. However, later reviewers have criticized many of these

studies because their small sample sizes and different methodologies make valid comparisons questionable.

In general, assessment programs are valid if their content is job-related. To the extent the dimensions resulted from an accurate and complete job analysis and to the extent that the selected exercises and procedures accurately measured dimensions, the procedure is valid. Of 22 published research studies attempting to evaluate the overall validity of assessment center applications, the majority (15 studies) indicated positive results for validity. On the whole, the studies were of extremely high quality and scientific rigor, especially compared to research on other management selection or development programs. As an example, a study of AT&T Salesmen Selection based on uncontaminated assessment center data yielded:

$r = .51$ for global rating from assessment panel and field review rating

$r = .33$ when only four paper and pencil tests were used against the criteria

In general, assessment center validity research during the last 10 years has not kept pace with earlier research. However, there are some new results available, including 20-year follow-up data on assessees. The data indicate that the assessment center is a viable predictor of management success, even as high as four levels above entry-level management positions (Byham, 1980).

Most operational validity studies are flawed by potential contamination since assessment data are used by the organization doing the evaluation. However, results that are available generally indicate that assessment centers are better at predicting ratings of management potential and actual advancement than they are at predicting performance at current jobs. The following are typical results of operational studies:

IBM - Eleven studies of its assessment center programs all show a significant positive relationship between center findings and various criteria of success.

Sears Roebuck - Most components of their assessment center were found to be significantly related to various criteria of job success.

Byham's 1971 article noted few studies that explored possible differential impacts of race or sex on assessment center scores. Current literature addresses this issue; in general, the assessment center is found to be equally fair for both sexes and for all races. While there were no court cases dealing with assessment centers up to 1970, there have been several in the last decade. No court has found assessment centers to be illegal or unfair. There have been several cases where assessment centers have gained the official blessing of a court as part of an affirmative-action effort to overcome past discrimination. It is also interesting to note that assessment centers are now being used by the EEOC itself to fill top management positions.

A summary table of validity information is displayed in Table 1. In addition to the summary table, a detailed analysis of validity information provided in reports from 27 studies is presented in appendixes. These 27 studies were divided into two groups: 14 related to overall effects of assessment (Appendix A) and 13 related to specific assessment exercises (Appendix B). The following information was extracted: identification of study, brief overall abstract, type of assessment or element represented, management level assessed, and validities reported, including definition as predictive, concurrent, or other. The 27 studies are listed by category, and in chronological sequence to highlight the changes that have taken place in assessment over time.

ASSESSOR LEVEL AND QUALIFICATIONS

Assessor Level

The appropriate level of those doing an assessment is not addressed directly in the literature. This section will present what is available and relate it where possible to studies that identify assessors at least by profession or degree. Where such identification of professional status is provided, it is possible to infer the level of training and to suggest adjustments in professional caliber to improve the quality and validity of the assessment process.

The consensus of professional psychologists working in the assessment field is that only professional psychologists with appropriate background and specialized training should administer and interpret the results of psychometric testing, and particularly the results of projective testing. At the least, the feedback interview and/or utilization of the information for decisions regarding individual careers should include a psychologist with professional training and experience in executive assessment.

In industrial settings, it appears that assessors are usually line managers two or more levels above the participants. Line managers are used because:

- they are familiar with the jobs
- participation as an assessor is a developmental experience
- management involvement increases program acceptance
- participation increases familiarity with the program and assures more effective use of results

The last factor is extremely important at a later stage of the process, since the manager will know the basis for assessment observations and judgments and can weigh them against job performance and other information.

A few organizations mix line and personnel department or other staff members as assessors, usually because of difficulty in recruiting assessors or to reduce costs. Professional psychologists are more often used when

Table 1

Validities and Levels by Assessment Program

Reference	Type of Assessment	Level of Management*	Validity (concurrent or predictive)
Hilton (1955)	psychometric and interviews	2 to 3	predictive $r = .29 - .38$
Laurent (1962)	psychometric and interviews	2 to 3	predictive moderately positive
Huse (1962)	psychometric, projective tests, interviews, final ratings	2 to 3	predictive $r = .07 - .44$ ave. $r = .19$
Thompson (1970)	psychometric, projective tests, simulation exercise, interviews, oral presentation	2 to 4	predictive $r = .38 - .85$
Campbell (1962)	social skills scales, creativeness, and overall rating	2 to 3	concurrent $r = -.05 - .50$
Albrecht (1964)	interview, psychometric, projective, human relations problems	4 to 5	predictive .34 to .46
Dicken (1965)	psychometric	2 and 3	predictive satisfactory
Bray (1968)	psychometric tests, interview, management exercises	2	predictive overall $r = .51$ paper-pencil tests: median $r = .33$
Kraut (1972)	psychometric tests, interview, situational exercises; developmental exercises	2 to 3	predictive: sig. differentiation of management potential
McConnell (1972)	management simulation exercises, interview	2 to 3	concurrent validity = .57
Ginsburg (1972)	psychometric tests, interview 2 situational exercises	2 to 3	predictive $r = -.34 - .37$ (most $-.30$'s)

Table 1 (Cont.)

Validities and Levels by Assessment Program

Reference	Type of Assessment	Level of Management*	Validity (concurrent or predictive)
Bray (1966)	psychometric tests, situational exercises, projective tests, interview, other	2 to 4	predictive .29 - .42
Byham (1971)	psychometric tests, situational exercises, interview	2 to 5	predictive .33 - .51
Cohen (1980)	pre-packaged programs, exercises	2 to 4	acceptable
Souder (1983)	a) traditional: psychometric tests, situational tests; interview	2 to 5	a) acceptable
	b) contractor operated	2 to 5	b) probably lower
	c) community center	2 to 5	c) where customized acceptable; where standardized, lower
	d) second generation centers	2 to 5	d) questionable
Townsend (1985)	psychometric, projective tests, simulated situational exercise	3 to 5	informal feedback indicates highly satisfactory predictive validity

*From Jacobs and Jacques (1985).

Level	Nature
7	- CEO
6	- Exec VP
5	- President
4	- Gen. Mgr.
3	- Unit Mgr.
2	- First Line Mgr.
1	- Direct Output

evaluation is in the very high levels of management. However, limited research indicates that professionals generally do no better than trained line managers. While professional psychologists may have superior observational skills, they lack company knowledge.

Assessor Qualifications

Finkle (in Dunnette, 1983) presents a discussion of managerial assessment centers that includes pieces of information useful for relating skills of assessors to the nature of the assessment process used. Finkle indicates that the backgrounds and qualifications of assessors vary from one program to another, but the use of a team of assessors is pervasive. This may be due to the popularity of small group situational exercises. Earlier assessments were more of the one-on-one variety. However, the same functions are performed by one assessor or by a team; the role of the assessor has only changed in degree.

Assessor training and requisite skill levels for valid administration and interpretation of test results vary with the nature and type of assessment instrument. One rule of thumb is that the more objective the test situation (the more it involves psychometric data such as percentiles in relation to particular group norms or the administration and interpretation of projective tests), the more a trained psychologist is needed for interpretation of results. However, as the test situation moves more toward the subjective end of the scale, e.g., situational tests and leaderless group discussions, the experience base of a managerial assessor provides greater competence to interpret individual performance. Even then, the nonprofessional assessor should be knowledgeable concerning the effects of stereotyping, leniency, halo, and central tendency errors on ratings. If projective tests are included as part of the assessment exercises, including a skilled psychologist on the assessment team provides useful confidence building for the manager-assessor and contributes to the acceptance of the assessment results.

Many organizations that are in the process of starting assessment centers initially train large numbers of managers to be members of an assessor "pool." Benefits include providing an orientation to the program to a large number of the managers who will ultimately use the assessment reports. In most centers, assessors serve only once, but management assessors in smaller companies are used more often. Training for manager-assessors is usually from 2-5 days. Training is seen as important, both for accurate assessment results and for the benefits to the assessors themselves.

Attending an assessment center is reported as the best way to understand the process. Many organizations have arranged for key managers to attend a center run by another company as a way of persuading management to adopt the technique. Showing videotaped recordings of assessment centers in operation is used as an alternative. Another effective means of acquainting managers with these methods is to put them through a representative exercise so that they experience the kinds of behaviors elicited by the exercises.

RELATIVE COSTS/BENEFITS

In today's budget-conscious military, an important issue is whether or not formal assessment programs for the identification and development of executive talent will pay off in dollars and cents. The literature survey investigated (1) practical and theoretical cost variables, (2) start-up costs, (3) benefits, and (4) a potential formula to calculate cost/benefits in quantitative terms. The following section contains a review of the relative costs and benefits associated with assessment, based on the available literature. A summary table integrating these findings is included.

Cost Variables

The literature on costs/benefits of assessment reflects more agreement on the predictors of success than on criteria that define success. Predictors are usually psychometric tests or subjective observations. The absence of consensus on criteria appears to be the result of problems of unreliability and subjectivity which affect the predictors less strongly. Unfortunately, criterion reliability impacts on predictor validity, and thus on cost/benefit considerations.

The problem can be illustrated by example. Assessments are not perfectly reliable and are not perfectly correlated to criteria of performance. The actual correlations between the two are generally in the low to middle range. The common variance between the two sets of measures (the square of the correlation coefficient) determines the potential benefits to be gained from assessment in relation to the costs incurred.

An additional problem is the complexity of the executive/organizational criteria. Because of this complexity, assessment variables must be carefully chosen if they are to account for much of the variance. Thus, for most assessment tasks, the major problem is in developing valid, meaningful criteria for the assessments. The simpler and more measurable a criterion, the more accurately it generally can be predicted from assessment data. Thus, estimates of best cost/benefit outcomes will be facilitated by selection of reliable predictors and a criterion of success which is quantifiable and reliable.

Speaking to this problem, Laughlin and Kedzie (1980) indicated that attempts to justify costs of evaluation, development, and follow-up of higher level executives probably will not be successful because the number of external variables is so great. They assert that the strongest case for executive developmental efforts is that in their absence, there will be human and financial losses resulting from the older, historically-condoned evaluation activities. While this is a positive argument for favorable cost/benefit outcomes, it is not a quantifiable outcome.

Ginsburg and Silverman (1972) described a one-day assessment program which appears to be a significant and relatively inexpensive means for inventorying human resources within an organization. However, the authors present attitudes rather than a listing of variables that could be included as metrics in a calculation of assessment cost/benefit outcomes.

Souder and Leksich (1983) pointed out that a major factor in the effectiveness of an assessment center (and thus its payoff) is the assessors and their training. It is to be noted also that assessors are one of the major items of expense in an assessment center. Contributing to assessor costs are: (1) time involved in the assessor training, (2) ratio of assessors to participants, (3) objectives of the center, which affects assessor skill requirements, (4) total number of participants, and (5) types of exercises, i.e., tests, situations and techniques used. An attempt to estimate the cost/benefits of an assessment program would need to include these variables and the metrics that can be derived from them.

Bender's (1973) research presented categories of cost factors usually encountered by assessment centers across the country. Although cost figures are from 1973, it seems unlikely that they have changed materially in terms of percentages since that time. Direct cost factors included (1) primary evaluation devices used in assessment, (2) operating procedures followed in center administration, and (3) level of assessee-assessor training. Indirect measurements are made from (1) center operating characteristics, (2) evaluation parameters most frequently observed, and (3) uses made of performance data, and the resulting documentation required. These categories could be valuable in anticipating the parameters of cost typically encountered.

As shown in the literature, the cost per subject in a "traditional" executive-potential assessment programs varies widely. Factors to include in calculating costs include (1) assessors' time, including training; (2) participants' time; (3) administrators' time, including preparing for the center and writing reports; (4) cost of meals and facilities; (5) cost of exercises (usually one-time investment with reusables or development of unique exercises); (6) consulting help; and (7) start-up costs.

Start-Up Costs

Byham (1980) dealt specifically with initiating an assessment situation in an update of the information in his earlier publication. A list of steps involved in starting an assessment center provided an indication of potential costs:

1. Determine objectives of program
2. Define dimensions to be assessed
3. Select exercises that will bring out the dimensions
4. Design assessor training and assessment center program
5. Announce program, inform participants and assessors, handle administrative detail
6. Train assessors
7. Conduct center
8. Write summary reports on participants
9. Feedback to participants a summary of performance at center and development actions
10. Evaluate center
11. Set up procedures to validate center against a criterion of job success

A major factor in start-up costs is the organization's need for outside consultants. Many organizations take information from reports such as Byham's, order exercises, and start the assessment center. Other organizations send their potential assessment center administrators to workshops on assessor training. Many others use consultants to aid in planning, assessor training, administration of several pilot programs, the initial writing of assessment center reports, and the planning of feedback interviews. Consultants can make their greatest contribution in planning a center and in assessor training.

Benefits

Benefits to be achieved by an assessment program are assumed to be synonymous with the goals of the program. The major difficulty in establishing criteria for goals is that evaluators often confuse system effectiveness with individual effectiveness. At the least, they often measure the one, hoping thereby to measure the other. This can be labeled a proximal-distal problem.

In an assessment situation, comparing the costs of an individual's assessment with the benefits to be derived by the organization to which the individual makes contributions is a distal problem of measurement. It raises the problem of separating out the contribution and/or hindrance of the individual to the organization when all other sources of contribution and/or hindrance are removed or controlled. This is virtually an impossible task. If the individual's performance is measured intermixed with all other contributing factors, e.g., as improvement in some organizational criterion, then the system's effectiveness has been measured, but not that of the individual.

The proximal problem or "solution" consists of measuring individual performance as close to the individual as possible and with as little influence as possible from other factors that might determine his/her performance. Thus, performance on tests and in test situations where the individual is either the sole performer or where his/her performance can be directly measured is desirable. In the proximal philosophy, one must be satisfied with upping the probability that future success will occur, rather than with actually measuring the individual's contribution to the organization's success. If one moves farther out toward the distal end of the range and attempts to measure benefits related to the individual's contribution, the numbers become less defensible.

A factor to be considered in determining benefit criteria is individual benefits that extend to not only participants but also to assessors. Managers who are trained to act as assessors in an assessment center have been found to benefit by:

- improved interviewing skills
- broadened skills in observation
- increased appreciation of group dynamics and differential leadership styles
- new insights into behavior

- strengthening management skills through practice of exercises
- broadened repertory of responses to problems
- establishing normative standards for performance evaluations
- more precise vocabulary to describe behavior

However, the greatest impact has been found in manager performance, especially appraisal interviewing. Assessor training provides a unique opportunity for managers to focus on observing behavior without the normal interruptions associated with business. Further, the principal focus of assessment training is usually on (1) interviewing, (2) observing behavior; and (3) handling the in-basket, skills that the assessors need in their own work.

Another benefit of assessment centers, especially when combined or associated with training, is the "unfreezing" process; that is, participants (and manager/assessors) are sensitized to their own shortcomings and are open to development ideas and training. A benefit of training following assessment is the correction of common deficiencies such as (1) inefficient group procedures, (2) public speaking, (3) insensitivity to others, (4) management skills, and (5) decision-making.

Souder and Leksich (1973) found improved interpersonal understandings and abilities in those who go through an assessment center. Assessor training also provides an opportunity to sharpen thinking and general perceptions. A further benefit is the stimulation of more effective performance appraisals and career path efforts. However, they caution that for these benefits to be achieved certain requirements must be met. As they state:

A system perspective must be taken in designing the assessment center. The center must be an integral part of an organization's performance appraisal, training, career path and counseling systems. The center must be consistent with the organization's established policies, goals and culture.

The center must have top management support and involvement. Top management must provide the guiding objectives and the criteria for judging the effectiveness of the center. Top management must provide feedback on how well the center is meeting these criteria.

The participants should be selected carefully, according to some pre-specified selection criteria that relate to the center's objectives. Not everyone should attend the assessment center, and supervisors should be instructed carefully on the types and categories of candidates desired. In that regard, it is important that some superior performers and a variety of 'problem children' be included in the participants. This provides both a healthy variety and a validity check on the center's procedures.

The credibility of the center must be maximized by maintaining valid exercises, competent assessors and honest feedback to the participants. Validity checks should be run constantly on the exercises, since they can become outdated. Participants should be polled to gather information on the credibility of the center.

The center should be viewed and managed as an agency of change. Tomorrow's managers, tomorrow's culture and tomorrow's values can come from the center. But in order to achieve this, the center must be properly focused and administered. If the center is too pedantic, it will simply reinforce the existing organizational values and climates. But if the center is too radical, it won't be accepted.

Quantitative Formula

A formula for calculating cost/benefits of assessment practices was not found in the literature. However, there was an attempt by Sauter (1980) to calculate the cost/benefit of Federal Executive Institute (FEI) programs, especially in the area of interpersonal effectiveness. His methodology involved determining the value to the agency of development training by asking for the impressions of both the executive and the supervisor responsible for sending him/her to FEI. The benefits to the agency were measured against the expenditure of the agency. Benefit was indicated by estimating any change in the executive's value to the agency as a result of the program. The tabled results follow:

	Executive estimate	Supervisor estimate
Average benefit/executive	\$101,808.00	\$25,588.00
Average cost/executive	10,907.00	10,215.00
Average net benefit/executive	90,901.00	15,373.00
Ratio benefits to cost	9.3	2.5

This simplistic approach is not seen as either valid or reliable. However, the literature suggests by the absence of other approaches that this approach or other subjective (and therefore equally suspect) approaches are the only ones available at the present time. Ways to measure the impact of assessment on organizational mission performance must be developed and are under current study.

Cohen (1980) lists cost/benefit issues related to the decision to use a pre-packaged assessment program. He lists the following considerations: (1) such programs may be more cost/effective if a careful in-house job analysis matches needs with existing assessment center materials, (2) they may be cost effective if used on a trial basis prior to developing one's own assessment program, and, (3) they may be cost effective if the intention is to assess only a few candidates, thereby avoiding start-up costs for an in-house center (but there needs to be a good match).

As noted before, the most difficult requirement for a cost/benefit equation is the criteria that would serve as the goal of the assessment system. The goal of assessment would determine the assessment center design required to reach the goal. The assessment system would then be translatable into processes such as tests and procedures consistent with the most valid and relevant information contained in the literature. Having established the system and identified its elements, the pricing of these elements according to the type of assessments conducted would not be difficult. On the other hand, the translation of the goal of the system into dollars and cents would be. If agreement could be reached on an amount and was consistent with the literature reported in this study, then the writing of an algorithm to relate costs to benefits could be accomplished. An optimizing strategy could be defined using a linear programming model, with various degrees of payoff for different levels of expenditure.

A summary table showing cost/benefit analyses across several studies is presented in Table 2.

SUMMARY OF LITERATURE SURVEY FINDINGS

This report has presented a literature survey of assessment technology related to the evaluation of industry executives. The information was gathered to assist the Army in defining appropriate assessment technology for use at approximately the 10- and 20-year career points for Army officers, and civilians.

There was a lack of specific information in the literature concerning 10- and 20-year career points. This probably reflects a concentrated interest on the potential of young, low- and mid-level managers for future high-level positions. Research is conducted, in the main, on department- and division-level individuals who are at the direct level of management. Literature directed at executive levels may be in sources not included in this literature survey.

The background literature published by experts in the assessment field generally supports trait theory, rather than a process- or situation-oriented construct, as the theoretical underpinnings for assessment. However, in reviewing actual assessment programs, it is clear that these procedures are moving more toward exercises and job simulation than toward classical psychometric testing. Current literature indicates that the psychometric test-based assessment with very little situational testing of a decade ago has now been reversed. Assessment theorists appear willing to accept a combination of trait and situationism if both are consistent with the goals of the assessment being conducted.

Byham (1980) provided a valuable summary and "report card" on current assessment technology. He estimates the number of companies operating assessment centers at 2,000, and the number of people evaluated by

Table 2

Summary Table of Assessment Level, Validity, Purpose, Costs, and Assessor Level

Reference	Type of Assessment	Level of Management*	Validity (concurrent or predictive)	Cost	Purpose	Level of Assessor
Hilton (1955)	psychometric and interviews	2 to 3	predictive $r = .29 - .38$	\$250/sub	prediction of job success, development needs	psychologists
Laurent (1962)	psychometric and interviews	2 to 3	predictive moderately positive	\$130/sub	management potential	psychologists, trained staff
Huse (1962)	psychometric, projective tests, interviews, final ratings	2 to 3	predictive $r = .07 - .44$ ave. $r = .19$	\$255/sub	prediction of job performance	psychologists
Thompson (1970)	psychometric, projective tests, simulation exercise, interviews, oral presentation	2 to 4	predictive $r = .38 - .85$	\$365/sub	managerial potential	staff psychologists, manager-assessors
Campbell (1962)	social skills scales, creativeness, and overall rating	2 to 3	concurrent $r = -.05 - .50$	\$60/sub	prediction of job performance	psychologists
Albrecht (1964)	interview, psychometric, projective, human relations problems	4 to 5	predictive .34 to .46	\$130/sub	managerial and executive potential	professional consultant firm
Dicken (1965)	psychometric	2 and 3	predictive satisfactory	\$150/sub	validity study; work performance	psychologists
Bray (1968)	psychometric tests, interview, management exercises	2	predictive overall $r = .51$ paper-pencil tests: median $r = .33$	\$50/sub	selection and job performance	trained mgrs
Kraut (1972)	psychometric tests, interview, situational exercises; development mental exercises	2 to 3	predictive: sig. differentiation of management potential	\$220/sub	validity study; managerial potential	psychologist, trained staff
McConnell (1972)	management simulation exercises, interview	2 to 3	concurrent validity = .57	\$80/sub	managerial potential and developmental evaluation	trained assessors (24-28 hours)
Ginsburg (1972)	psychometric tests, interview 2 situational exercises	2 to 3	predictive $r = -.34 - .37$ (most -.30's)	\$25/sub	skill and ability assessment; developmental needs	trained assessors

Table 2 (cont.)

Summary Table of Assessment Level, Validity, Purpose, Costs, and Assessor Level

Reference	Type of Assessment	Level of Management*	Validity		Cost	Purpose	Level of Assessor
			(concurrent or predictive)				
Bray (1966)	psychometric tests, situational exercises, projective tests, interview, other	2 to 4	predictive .29 - .42		\$275/sub	managerial (success) potential; developmental potential	psychologists, "professionally trained persons"
Byham (1971)	psychometric tests, situational exercises, interview	2 to 5	predictive .33 - .51		varies \$100-300/sub	managerial potential	psychologist plus trained staff
Cohen (1980)	pre-packaged programs and exercises	2 to 4	acceptable		\$50-150/sub	prediction of job performance; developmental needs	trained assessors
Souder (1983)	a) traditional: psychometric tests, situational tests; interview	2 to 5	a) acceptable		Costs: -assessor training -customizing -program -ratio of assessors to participants -total number of participants -types of tests	management potential, developmental needs	a) psychologist and trained staff
	b) contractor operated	2 to 5	b) probably lower				b) trained staff
	c) community center	2 to 5	c) if customized, acceptable; if standardized, lower				c) trained staff
	d) second generation centers	2 to 5	d) questionable				
Townsend (1985)	psychometric, projective tests, simulated situational exercise	3 to 5	informal feedback indicates highly satisfactory predictive validity		\$485/sub	hiring, promotion, executive potential	1-PhD, 1-M.A., technician

*From Jacobs and Jacques (1985):

Level	Nature
7	- CEO
6	- Exec VP
5	- President
4	- Gen. Mgr.
3	- Unit Mgr.
2	- First Line Mgr.
1	- Direct Output

American Telephone and Telegraph alone at 200,000. Also, job analysis methodologies that document the precise dimensions and exercises used in the assessment center have become much more sophisticated. Assessment center procedures are now more behavioral and descriptive than 10 years ago. Assessments have also become shorter because of improved assessor training technology and clearer observation forms.

An important issue is the validity of the procedures that make up an assessment battery. The overlap between the assessment procedures used by different organizations was so great that it was not possible to separate them into categories of validities for specific types of assessment. Each of the assessment programs reviewed had a common core of psychometrics with varying other assessment procedures idiosyncratic to the organization. Analyses related the different combinations of assessment procedures to their respective validities, cost/benefit information, level of personnel assessed, purpose of the assessment, and skill level of the assessor.

The literature is clear that satisfactory validities are available for the various tests, techniques, and procedures that would most likely form part of an assessment program to be established by the Army. Any test or procedure has as many different reliabilities and validities as there are different situations in which it is applied. An Army assessment program constructed of elements reported as having satisfactory validities will still need its own validity-testing as a new program. However, the earlier evaluative research will significantly shorten the time required to establish a program, and increase the likelihood of its satisfactory statistical support and, subsequently, its credibility.

The skill levels required of assessors in different types of evaluations varied from Ph.D. psychologists down to company personnel trained as lay managers-assessors. Some manager-assessors received extensive training, such as a three-week course at a psychological institute. Others received only company training, sometimes from a staff psychologist and sometimes from vague training courses.

In the area of cost/benefits, the literature contained less information than is needed for sound estimates of the cost of various programs with differing validities. Most of the cost/benefit information appeared in the form of descriptive phrases such as "expensive" or "reasonable cost." The search for cost/benefit information addressed the following: metrics relating cost to benefits under certain assessment procedures; a list of variables and their quantification that could be used in such a formula; indications of the actual costs incurred under certain assessment conditions; and subjective information reflecting the opinions of researchers who had evaluated the costs/benefit of different assessment conditions.

The most difficult criteria to quantify are the goals and objectives of assessment and development programs. Individual development can be measured by comparing test scores and performance ratings longitudinally. Performance measurements, however, are confounded by organizational variables that either assist or hinder an individual in his/her work.

Estimating costs and benefits for an Army program must be deferred until the goals and components of assessment are established, but some of the relevant variables to include in such a formula are presented. While no generic formula was found, the literature survey did confirm that it is possible to define the metrics of assessment in relationship to cost/benefit calculations for a particular assessment situation.

REFERENCES

- Albrecht, P., Glaser, E., & Marks J. (1964). Validation of a multiple-assessment procedure for managerial personnel. Journal of Applied Psychology. 48(6), 351-360.
- Bender, J. (1973). What is "typical" of assessment centers. Personnel Journal. 50(4), 50-57.
- Blake, R. & Mouton, J. (1982). Theory and research for developing a science of leadership. Journal of Applied Behavior Science. 18(3), 275-291.
- Bowers, K. (1973). Situationalism in psychology: An analysis and a critique. Psychological Review. 80(5), 307-336.
- Bray, D. & Campbell, R. (1968). Selection of salesmen by means of an assessment center. Journal of Applied Psychology. 52(1), 36-41.
- Bray, D. & Grant, D. (1966). The assessment center in the measurement of potential for business management. Psychological Monographs. 80(17), 1-27.
- Byham, W. (1971). Validity of assessment centers. Proceedings, 79th Annual Convention, American Psychological Association.
- Byham, W. (1980). The assessment center as an aid in management development. Training and Development Journal. 34(6), 24-36.
- Campbell, J., Otis, J., Liske, R., & Prien, E. (1962). Assessments of higher-level personnel: II. Validity of the overall assessment process. Personnel Psychology. 15(1), 63-74.
- Carlton, F. (1970). Relationships between follow-up evaluations and information developed in a management assessment center. Proceedings, 78th Annual Convention, American Psychological Association, 565-566.
- Cohen, S. (1980). Pre-packaged vs. tailor-made: The assessment center debate. Personnel Journal. 59(12), 989-991.
- Comrey, A. & High, W. (1955). Validity of some ability and interest scores. Journal of Applied Psychology. 39(4), 247-248.
- Dicken, C. & Black, D. (1965). Predictive validity of psychometric evaluations of supervisors. Journal of Applied Psychology. 49(1), 34-47.
- Dodd, W. (1970). Will management assessment centers issue selection of the same old types? Proceedings, 78th Annual Convention, APA, 569-570.
- Finkle, R. (1983). Managerial assessment centers, in Handbook of Industrial and Organizational Psychology. New York: John Wiley & Sons, 861-888.

- Finley, R. Jr. (1970). Evaluation of behavior predictions from projective tests given in a management assessment center. Proceedings, 78th Annual Convention, APA, 567-568.
- Frederickson, N., Saunders, D. & Wand, B. (1957). The in-basket test. Psychological Monographs. 71(9), 1-28.
- Ginsburg, L. & Silverman, A. (1972). The leaders of tomorrow: Their identification and development. Personnel Journal. 51(9), 662-666.
- Grant, D. (1965). Analysis of management potential. Presented at the 1965 Meeting of the Eastern Psychological Association.
- Grant, D. & Bray, D. (1969). Contributions of the interview to assessment of management potential. Journal of Applied Psychology. 53(1), 24-34.
- Grant, D., Katkovsky, W. & Bray, D. (1967). Contributions of projective techniques to assessment of management potential. Journal of Applied Psychology. 51(3), 226-232.
- Greenwood, J. & McNamara, W. (1967). Interrater reliability in situational tests. Journal of Applied Psychology. 51(2), 101-106.
- Guilford, J. (1959). Personality. New York: McGraw-Hill.
- Haggerty, H. (1953). Personnel research for the U.S. Military Academy 1942-1953. U.S. Army Personnel Research Board Technical Research Report 1077.
- Haggerty, H. (1963). Status report on research for the U.S. Military Academy (Cadet Leaders Task). U.S. Army Personnel Research Office.
- Handyside, J. & Duncan, D. (1954). Four years later: A follow-up of an experiment in selecting supervisors. Occupational Psychology. 28, 9-23.
- Hardesty, D. & Jones, W. (1968). Characteristics of judged high potential management personnel: The operations of an industrial assessment center. Personnel Psychology. 21(1), 85-98.
- Hinrichs, J. (1969). Comparison of 'real life' assessments of managerial potential with situational exercises, paper-and-pencil ability tests, and personality inventories. Journal of Applied Psychology. 53(5), 425-432.
- Hilton, A., Bolin, S., Parker, J. Jr., Taylor, E. & Walker, W. (1955). The validity of personnel assessments by professional psychologists. Journal of Applied Psychology. 39(4), 287-293.
- Hollander, E. (1965). Validity of peer nominations in predicting a distant performance criterion. Journal of Applied Psychology. 49(6), 434-438.

- Huse, E. (1962). Assessments of higher-level personnel: IV. The validity of assessment techniques based on systematically varied information. Personnel Psychology. 15(2), 195-205.
- Jacobs, T.O. and Jacques, E. (1985). Leadership in complex systems. In: Human productivity enhancement, Vol. II: Organization and personnel, J. A. Zeidner (ed.). New York: Praeger.
- Kirkpatrick, J. (1966). Five-year follow-up study of the American Chamber of Commerce standard application for employment form. American Chamber of Commerce.
- Knauff, E. (1949). A selection battery for bake shop managers. Journal of Applied Psychology. 33(4), 304-315.
- Korman, A. (1968). The prediction of managerial performance: A review. Personnel Psychology. 21(3), 295-322.
- Kotula, L. and Haggerty, H. (1966). Research in the selection of officer candidates and cadets. U.S. Army Personnel Research Office.
- Krauss, I. (1964). The predictive value of certain tests for company advancement: A follow-up study. Unpublished M.A. thesis, New York University.
- Kraut, A. & Scott, G. (1972). Validity of an operational management assessment program. Journal of Applied Psychology. 56(2), 124-129.
- La Gaipa, J. (1960). Development of officer candidate biographical information blank: I. Validation of some experimental inventories. Bureau of Naval Personnel Technical Bulletin. 60-69.
- Laughlin, T. & Kedzie, D. (1980). Building a better executive development mousetrap. Training and Development Journal. 34(3), 80-83.
- Laurent, H. (1962). Early identification of managers. Management Record. 24(5), 33-38.
- Lawshe, C. (1949). How can we pick better supervisors? Personnel Psychology. 2, 69-73.
- MacKinney, A. & Wolins, L. (1960). Validity information exchange. Personnel Psychology. 13(4), 443-447.
- McConnell, J. & Parker, T. (1972). An assessment center program for multi-organizational use. Training and Development Journal. 26(3), 6-14.
- Meyer, H. (1974). The science of telling executives how they're doing. Fortune Magazine. 89(1), 102-112.
- Mischel, W. (1968). Personality and Assessment. New York: Wiley.

- Osborn, R., Lawrence R., Martin, T., & Glueck, W. (1981). The event of CEO succession, performance, and environmental conditions. Academy of Management Journal. 24(1), 183-191.
- Phelan, J. (1962). Projective techniques in the selection of management personnel. Journal of Projective Techniques. 26(1), 102-104.
- Porter, A. (1962). Personality ratings by faculty as predictors of executive success for graduate business school alumni. Occupational Psychology, 26, 1962, 146-151
- Ricciuti, H. (1955). Ratings of leadership potential at the U.S. Naval Academy and subsequent officer performance. Journal of Applied Psychology. 39, 194-199.
- Robbins, J. & King, D. (1961). Validity information exchange. Personnel Psychology. 13(2), 217-219.
- Sauter, J. (1980). Purchasing public sector executive development. Training and Development Journal. 35(2), 333-347.
- Scollay, R. (1957). Personal history data as a predictor of success. Personnel Psychology. 10(1), 23-26.
- Soar, R. (1956). Personal history data as a predictor of success in service station management. Journal of Applied Psychology. 40(6), 383-385.
- Souder, W. & Leksich, A. (1983). Assessment centers are evolving toward a bright future. Personnel Administrator. 28(11), 80-87.
- Thomson, H. (1970). Comparison of predictor and criterion judgments of managerial performance using the multitrait-multimethod approach. Journal of Applied Psychology. 54(6), 496-502.
- Thorndike, R. & Hagen, E. (1959). Ten thousand careers. New York: John Wiley & Sons.
- Townsend, J. & Alderson, M. (1985). Executive assessment program. Washington, D.C.: Catholic University.
- Ulrich, L. & Trumbo, D. (1965). The selection interview since 1949. Psychological Bulletin. 63(2), 100-116.
- Verlander, E. (1985). The system's the thing. Training and Development Journal. 39(4), 20-23.
- Williams, F. & Harrell, T. (1964). Predicting success in business. Journal of Applied Psychology. 48(3), 164-167.
- Williams, S. & Leavitt, H. (1947). Group opinion as a predictor of military leadership. Journal of Consulting Psychology. 11, 283-291.

Wollack, L. & Guttman, I. (1960). Prediction of OCS academic grades and post-OCS performance of junior officers with a battery of speeded tests. U.S. Naval Personnel Research Field Activity.

Wollowick, H. & McNamara, W. (1969). Relationship of the components of an assessment center to management success. Journal of Applied Psychology. 53(5), 348-352.

APPENDIX A

Index to Appendix A

The following studies are reviewed in Appendix A. They reflect validity data from research associated with assessment in general. They are presented in chronological order.

1. Hilton, A., Bolin, S., Parker, J. Jr., Taylor, E. & Walker, W. (1955). The validity of personnel assessments by professional psychologists. Journal of Applied Psychology. 39(4), 287-293.
2. Campbell, J., Otis, J., Liske, R., & Prien, E. (1962). Assessments of higher-level personnel: II. Validity of the overall assessment process. Personnel Psychology. 15(1), 63-74.
3. Huse, E. (1962). Assessments of higher-level personnel: IV. The validity of assessment techniques based on systematically varied information. Personnel Psychology. 15(2), 195-205.
4. Laurent, H. (1962) Early Identification of Managers. Management Record. 24(5), 33-38.
5. Albrecht, P., Glaser, E., & Marks J. (1964). Validation of a multiple-assessment procedure for managerial personnel. Journal of Applied Psychology. 48(6), 351-360.
6. Dicken, C. & Black, D. (1965). Predictive validity of psychometric evaluations of supervisors. Journal of Applied Psychology. 49(1), 34-47.
7. Bray, D. & Campbell, R. (1968). Selection of salesmen by means of an assessment center. Journal of Applied Psychology. 52(1), 36-41.
8. Korman, A. (1968). The prediction of managerial performance: A review. Personnel Psychology. 21(3), 295-322.
9. Thomson, H. (1970). Comparison of predictor and criterion judgments of managerial performance using the multitrait-multimethod approach. Journal of Applied Psychology. 54(6), 496-502.
10. McConnell, J. & Parker, T. (1972). An assessment center program for multi-organizational use. Training and Development Journal. 26(3), 6-14.
11. Kraut, A. & Scott, G. (1972). Validity of an operational management assessment program. Journal of Applied Psychology. 56(2), 124-129.
12. Byham, W. (1980). The assessment center as an aid in management development. Training and Development Journal. 34(6), 24-36.
13. Cohen, S. (1980). Pre-packaged vs. tailor-made: The assessment center debate. Personnel Journal. 59(12), 989-991.
14. Souder, W. & Leksich, A. (1983). Assessment centers are evolving toward a bright future. Personnel Administrator. 28(11), 80-87.

APPENDIX A

Validity Studies Associated with Assessment in General

1. Hilton, A., Bolin, S., Parker, J. Jr., Taylor, E. & Walker, W. (1955). The validity of personnel assessments by professional psychologists. Journal of Applied Psychology. 39(4), 287-293.

This was an early pilot study to see how well job areas can be evaluated independently through an assessment technique and to evaluate the reduction of descriptive verbal material to summary quantitative form so as to facilitate statistical analysis. The conclusions, in light of the limitations presented:

- 1) Compared with most validity findings, these results are promising and indicate the assessment technique investigated has practical value and worthy of further research
- 2) When research is more rigorously conducted, the resulting estimates of validity are likely to be higher.

Type of assessment: Prediction of job success and recommendations for future employee development.

Tests given: Cardall Practical Judgment Test, How Supervise

- Form A, Allport-Vernon, Strong, Guilford-Zimmerman Temperament, PRI Classification and Tabulation Tests
- Projective techniques (unidentified)
- Interview - by two psychologists

Management level: pre-management and low-management positions

Predictive validity: (N=100)

- For overall job success: .29
- For advancement potential: .38

2. Campbell, J., Otis, J., Liske, R., & Prien, E. (1962). Assessments of higher-level personnel: II. Validity of the overall assessment process. Personnel Psychology. 15(1), 63-74.

The purpose of the study was to investigate the validity of predictions made by the psychologist who prepared the final report for the client.

Predictive criteria: first- and second-level supervisory ratings of sales and non-sales personnel

Predictors: social skills scales, creativeness (intellectual functioning) and overall rating (N=143)

Correlation between Psychologists' and Supervisors' Ratings on
Identical Scales (concurrent validity)

<u>Rating Scale</u>	<u>1st Level Supervisor</u>		<u>2nd Level Supervisor</u>	
	<u>Sales</u>	<u>Non-Sales</u>	<u>Sales</u>	<u>Non-Sales</u>
Social Skills	.16	.10	.13	.24
Persuasiveness	.08	.22	.34	.31
Leadership	.35	.46	.33	.45
Intellectual Capacity	.14	.41	-.05	.34
Creativeness	.31	.40	.31	.50
Planning	.26	.12	.35	.18
Motivation and Energy	.39	.04	.19	.10
Overall Effectiveness	.13	.27	.17	.12

Author's conclusions:

- 1) Psychologists are able to make predictions of successful and unsuccessful job performance using a combination of interview information, objective tests, and clinical reports of projective test data.
 - 2) Industrial supervisors are more lenient in their assessment of individuals than industrial psychologists tend to be.
 - 3) It appears that prediction of a criterion rating worded in action terms is more effective than prediction of a criterion worded in behavioral terms.
3. Huse, E. (1962). Assessments of higher-level personnel: IV. The validity of assessment techniques based on systematically varied information. Personnel Psychology. 15(2), 195-205.

This study investigated changes in the validity of an assessment program as additional measurements were included. Concurrent validities were established between sets of ratings based on (a) complete assessment information made available to psychologists writing the reports (psychometric and projective tests, interviews, final ratings), (b) sub-elements of complete information, and, (correlated with) concurrent performance ratings of the assessee by a first level supervisor on eight appraisal dimensions.

Product Moment Validities of Predictor Raters Vs. Criterion Rating
(N = 107)

<u>Rating Scale</u>	<u>Interviewer Rating</u>	<u>Projective Rating</u>	<u>Test Rating</u>	<u>Final Rating</u>	<u>Report Rating</u>
1. Social Skills	.12	.18	.24*	.13	.13
2. Persuasiveness	.11	.33**	.22*	.22*	.24*
3. Leadership	.38**	.26**	.15	.44**	.28**
4. Intellectual Capacity	.26**	.13	.35**	.32**	.32**
5. Creativeness	.27**	.17	.34**	.41**	.23**
6. Planning	.12	.18	.35**	.21**	.29**
7. Motivation and Energy	.08	.03	.29**	.17	.07
8. Over-all Effectiveness	.16	.21*	.28**	.28**	.11**
Mean r (z transformation)	.19	.19	.28	.28	.21

*Significant at 5% level.

**Significant at 1% level.

Conclusions:

- 1) Psychologists can make relatively reliable ratings based upon psychometric data alone.
 - 2) The relative validity of ratings based on psychometric data appears to be higher than ratings based upon the interview or projective tests.
 - 3) The relative validity of ratings based upon complete information does not, in general, increase over ratings based solely upon psychometric tests. Available data suggest that the psychologist writing the final report may be unduly influenced by his impressions.
 - 4) Ratings based solely on the final report showed a consistent drop in validity when compared to ratings made on the basis of complete information, suggesting that the psychologist writing the report has difficulty in summarizing and reporting all valid information.
 - 5) A factor analysis indicates that most of the agreement between predictor and criterion raters was based more on general impressions than on specific variables; differential validity was greatest in the areas of intellectual ability and energy expended on the job.
4. Laurent, H. (1962) Early Identification of Managers. Management Record. 24(5), 33-38.

The purpose of this Standard Oil (NJ) research project report was the early identification of managerial potential, usefulness of tests, questionnaires and other information. It measured varying degrees of success (no unsuccessful people in study).

Tests: Millers Analogy, Non-Verbal Reasoning, Guilford-Zimmerman Temperament Survey, Individual Background Survey (experimental), Self-Performance Report (tried and dropped), survey of management attitudes, TAT-type projective test, individual interview

Predictive criteria: 1) position level (adjusted), 2) salary history, (salary and age adjusted), 3) a single criterion index, heavily weighted with position level, managerial effectiveness and salary. (N = 443, from all functional areas of company.)

Results: Some moderate but useful relationships between some of the standardized tests and the criterion, however the experimental tests provided the best results. The best test battery they could put together would consist of: Individual Background Survey, management judgment test, temperament test (own key), mental abilities test and self-performance report.

Predicting the Top Third of the Success Criterion by Weighted Test Scores

If candidate's weighted test scores is in the	...his chances in 100 of being in the top third of the success criterion are:
highest 20% of scores	76
next 20% of scores	47
next 20% (middle)	27
next 20%	13
lowest 20%	4

5. Albrecht, P., Glaser, E., & Marks J. (1964). Validation of a multiple-assessment procedure for managerial personnel. Journal of Applied Psychology. 48(6), 351-360.

A multiple-assessment procedure employing a personal history form, an intensive interview, two objective intellectual aptitude tests, a sentence completion test, and a human relations problems test was used to predict the performance of 31 industrial managers all having similar job assignments. Predictions were made on the basis of a global, nonactuarial analysis of the objective and subjective data; four sets of criterion judgments were obtained on four variables: three different sets of rankings and one set of ratings. A multitrait - multimethod matrix was used in the analysis of the intercorrelations. Nine of the 12 validity coefficients involving ranking-type criteria were statistically significant. Of the four coefficients involving rating type criteria, none was significant.

Type of assessment: administered for selection of managers and executives at approximately 11 years service (executive potential)

Paper and pencil tests: Problem test (aptitude), Sentence Completion Test, Human Relations Problems Test, Watson-Glaser Critical Thinking; personal history forms
Interview - two hours (by consulting staff)

Management level: middle-management (District managers)

Predictive validity: (N=31) The criterion was later on-the-job performance in terms of overall effectiveness:

By superiors: .34 By peers: .56 By consultants: .46

6. Dicken, C. & Black, D. (1965). Predictive validity of psychometric evaluations of supervisors. Journal of Applied Psychology. 49(1), 34-47.

Thirty-one higher level employees in one firm and 26 in another were assessed by objective test batteries. Clinical interpretations of test data, test scores, and other predictors were analyzed with reference to criterion personality ratings and management decisions at a follow-up point of 3 1/2 years for the first sample and 7 years for the second. Predictive validity of test assessments was generally satisfactory in the first sample, although not pragmatically superior to that of certain objective data. Prediction was less satisfactory in the second sample, but more unique to test data. A matching study indicated some correspondence of test reports and criterion personality sketches in the second sample. Uninterpreted test scores were not generally valid except as measures of intelligence. Implications of the sample differences and of the method were discussed.

Type of assessment: Counseling and Testing Center (not an "assessment center")

Objective test battery: 1) Strong Vocational Interest Blank, 2) MMPI, 3) Otis Quick Scoring Mental Ability (Gamma), 4) and others, i.e., Clerical Test, Mechanical Comprehension, Practical Judgment, How Supervise Rating Variables (assessment dimensions):

- 1) Effective Intelligence
- 2) Personal Soundness
- 3) Drive and Ambition
- 4) Leadership and Dominance
- 5) Likableness
- 6) Responsibility and Conscientiousness
- 7) Ability to Cooperate
- 8) Estimate of Potential Functioning level

Predictor data:

- 1) Report rating
- 2) Adjective and Phrase Analysis
- 3) Test Scores, and
- 4) Test ratings

Criterion data:

- 1) Field ratings
- 2) Objective criteria, i.e., termination, salary, change in job level, etc.

NOTE: Small samples and non-typical tests make this study of doubtful utility.

7. Bray, D. & Campbell, R. (1968). Selection of salesmen by means of an assessment center. Journal of Applied Psychology. 52(1), 36-41.

Newly hired candidates for sales positions were evaluated by means of an assessment center consisting of paper-and-pencil tests, an interview, and individual and group simulations. Assessment staff judgments were compared with job performance some months later as evaluated by a special observational team. Assessment results were strongly related to this criterion. Supervisors' and trainers' ratings were not significantly related to job performance criterion nor to assessment results. The findings lend support to recent studies indicating the efficacy of the assessment center method in personnel selection.

Type of assessment: selection and short-term prediction of job performance, (two day program)

Paper and pencil tests: SCAT, Critical Thinking, Contemporary Affairs Test, Abstract Thinking Test; biographical information; a lengthy interview by trained management assessors one at fourth level, six at third level, with three weeks training by psychologists;

Simulation exercises: Leaderless Group Discussion; Oral Fact Finding Exercise (individual); Consulting Case (individual)

Managerial level: pre-management, new hires

Predictive validity: the criteria was overall assessment judgment compared with performance ratings some months later:
overall assessment judgment: .51
four paper and pencil tests only: .33

8. Korman, A. (1968). The prediction of managerial performance: A review. Personnel Psychology. 21(3), 295-322.

The purpose of this article is to review and critically evaluate the research literature pertaining to the usefulness of various procedures in the prediction of leadership behavior in formal organizations in a selection context. The research paradigm is predictive validity. Research studies are examined in the following categories:

1. Psychometric Prediction - Cognitive Ability Tests
2. Psychometric Prediction - Objective Personality and Interest Inventories
3. Psychometric Prediction - Leadership Ability Tests
4. Psychometric Prediction - Personal History Data

5. Judgmental Prediction: Executive Assessments
6. Judgmental Prediction: Peer Ratings
7. Judgmental Prediction: Superior and Faculty Ratings

Relevant conclusions:

1. "Judgmental" prediction methods, as exemplified particularly by executive assessment procedures and peer ratings, are generally better predictors than psychometric procedures, although allowance must be made for the generally small samples involved.
2. There are at least seven research studies in the literature that have used the executive assessment procedure, and the results have been fairly promising and fruitful. The studies cited are:
Albrecht, Glaser and Marks (1964), Grant's AT&T Management Progress Study (1965), Dicken and Black (1965), Phelan (1962), supporting projective test protocols in prediction; Handyside and Duncan (1954), supporting a managerial selection panel; Meyer (1956); and Vernon (1950).
3. All in all then, there is little reason for anyone to be contemptuous of "judgmental prediction" in management selection. While the Ns were often small, the correlations were consistently at a usefully high level and certainly higher than the results found by the actuarial prediction methods.

1. Summary of Studies--Psychometric Prediction: Cognitive Ability Tests
*Significant at .05 level. **Significant at .01 level.

1 The general guides to this and the following tables are as follows:

- 1) Large Roman numerals indicate the sample sizes, results, etc. for separate samples and studies within a given report.
- 2) Small letters and small Arabic numbers indicate separate instruments and results within a given study or sample within a given report.
- 3) In general, the name of a familiar test is spelled out completely the first time it appears and then a familiar abbreviation is used, e.g., Strong Vocational Interest Blank = SVIB, etc.

Investigators	Description of Sample and Study	N	Predictors	Criterion	Results
Williams & Leavitt (1947)	Marine Corps Officers	185	a) AGCT b) Mechanical Aptitude	Ratings	a) $r = .02$ b) $r = .01$
Lawshe (1949)	Production Supervisors	70	Purdue Adaptability	Turnover	Significant negative relationship between the two variables
Comments: No correlations computed					
Knauff (1949)	Bakery Shop Managers	33	Wonderlic	Objective criterion and rating composite	$r = .26$
Comments: t-ratio between 16 "highs" and 16 "lows" was sig. at .05 level					
Meyer (1956)	Supervisors Various samples	I. 142 II. 45 III. 68	I. Wonderlic II. AGCT III. Bennett test of mechanical comprehension	Overall ratings	I. $r = .27^{**}$ II. $r = .13$ III. $r = .29^{**}$
Comments: Possible criterion contamination (i.e., managers were free to use predictor judgment in making criterion ratings).					
Handyside & Duncan (1954)	Supervisors	44	a) Verbal abilities b) Non-verbal abilities	Promotion rate	a) $r = .52^{**}$ b) $r = .40^{*}$

1. Summary of Studies--Psychometric Prediction: Cognitive Ability Tests (Cont.)

Investigators	Description of Sample and Study	N	Predictors	Criteria	Results
Mackinney & Molins (1960)	Supervisors (3 overlapping groups)	I. 66 II. 66 III. 45	I. a) Otis Employment b) Bennett c) Minnesota Paper d) Minnesota Clerical II. a) Otis Employment b) SRA Mechanical Apt. c) Minnesota Clerical III. a) Otis Employment b) Minnesota Clerical	I & II. Level, tenure & rankings III. Suggestion plan activity by supervisors and subordinates	I. Otis positively related; Bennett negatively related. No significant relationships reported. II. No significant relationships reported III. $r = .27^*$ (for Minn. Clerical Numbers)
Comrey & High (1955)	Supervisors	214 to 227	California Test of Mental Maturity	Objective Performance Data	No significant correlations were found
Comments: Use of groups, rather than individual scores, might have obscured relationships.					
Thorndike & Hagen (1959)	Managers	1,055	18 Ability Tests	Self-described pay, mobility & achievement	Results essentially negative in nature
Williams & Harrell (1964)	Stanford MBAs--2 samples: 1) self-employed and family business and 2) employed by large organization	I. 196 II. 116	Thorndike (or Ohio) Test	I. Salary II. Administrative Level	I. $r = -.01$ II. $r = -.01$
Comments: Sample undoubtedly highly pre-selected on verbal ability basis.					
Robbins & King (1961)	Sales Managers (5 samples)	100 to 538	Wonderlic	a) Position level changes b) Termination	Highest r in the matrix was found to be .11.

I. Summary of Studies--Psychometric Prediction: Cognitive Ability Tests (Cont.)

Investigators	Description of Sample and Study	N	Predictors	Criterion	Results
Grant (1965)	Managers--4 samples	40 to 84	a) SCAT Verbal b) SCAT Total	Position level changes	Relationship significant in only 1 of 4 samples (N = 83); SCAT Verbal $r = .46^*$, SCAT Total $r = .37^*$
Albrecht, Glaser, & Marks (1964)	Managers	31	a) General Mental-Ability Test b) Watson-Glaser Critical Thinking	Ratings on four separate dimensions	1 of 8 r 's significant at .05 level
Phelan (1962)	Managers	94	a) ACE Test of Scholastic Aptitude (ACE) b) Bennett Test of Mechanical Comprehension (BTMC)	Performance ratings	No significant relationships found.
Laurent (1962)	Managers--2 samples	200+ per sample	a) Miller Analogies Test (MAT) b) Non-verbal ability test	Ratings	I. a) $r = .18^{**}$ b) $r = .29^{**}$ II. a) $r = .20^{**}$ b) $r = .26^{**}$
Ricciuti (1955)	Naval Officers	308	a) Verbal reasoning b) Mechanical comprehension c) Mathematics course grades	Ratings	Highest r in the matrix was .08.
Hollander (1965)	Naval Officers	639	a) Verbal reasoning b) Mechanical comprehension c) Various course grades	Ratings	Correlations ranged up to +.22, with several significant due to large N.

I. Summary of Studies--Psychometric Prediction: Cognitive Ability Tests (Cont.)

Investigators	Description of Sample and Study	N	Predictors	Criterion	Results
Dicken & Black (1965)	Manufacturing Managers	I. 31	a) Otis Employment b) General Clerical c) BTMC Minnesota Paper d) Form Board (MPFB)	Ratings & Objective Criterion	I. a) & b) predicted several objective and ratings criteria but not all. c) & d) were less effective
	Insurance Managers	II. 26	a) Otis Employment b) Minnesota Clerical--Names c) Minnesota Clerical--Numbers	Ratings & Objective Criterion	II. a) predicted an intelligence rating criterion b) predicted the intelligence rating and a salary criterion c) predicted Leadership Potential ratings

Comments: There was a possibility of criterion contamination in this study, according to authors.

Vernon (1950)	Civil-Service Managers (British) 2 samples	I. 147 II. 202	A variety of Intellectual Ratings Ability Tests	Ratings	No correlation in either sample above .22.
---------------	---	-------------------	--	---------	--

11. Summary of Studies--Psychometric Prediction: Objective Personality and Interest Inventories

Investigators	Description of Sample and Study	N	Predictors	Criterion	Results
Comrey & High (1955)	Production Supervisors	214 to 227	Six Kuder Preference Record-- Vocational Scales (KVPR)	Objective Performance Data	R's insignificant except one which was opposite to logical expectation.
MacKinney & Wolins (1960)	Supervisors--3 overlapping samples	I. 66 II. 66 III. 45	a) Guilford-Zimmerman' b) Strong Vocational Interest Blank (SVIB)	I&II. Level, tenure and rankings III. Suggestion plan activity by supervisors and subordinates	Inconsistent random patterns of significant r's
Krauss (1964)	Supervisors--4 samples	23 to 73	KVPR	Level changes	Computational scale predicted significantly in 2 of the 4 samples; No other scale predicted.
Knauff (1949)	Bakery Shop Managers	33	Jurgensen Classification Inventory (empirical key)	Objective criterion and rating composite	$r = .39^*$
Comments: Jurgensen Inventory was later withdrawn from market.					
Williams & Harrell (1964)	Stanford MBAs--2 samples 1) self- and family-employed 2) large organization	I. 196 II. 116	SVIB	I. Salary II. Administr. Level	I. Only M-F Scale significant ($r = .19^*$) II. Only Personal Scale significant ($r = .24^*$)

II. Summary of Studies--Psychometric Prediction: Objective Personality and Interest Inventories (Cont.)

Investigators	Description of Sample and Study	N	Predictors	Criterion	Results
Grant (1965)	Utility Company Managers-- 4 samples	40 to 79	a) Edwards Personal Preference Schedule (EPPS) b) Guilford-Martin Inventory of factors (GAMIN)	Position level changes	No consistent pattern appeared across the 4 samples for either instrument
Phelan (1962)	Middle-level Managers	94	Allport Ascendancy-Submission Test	Performance ratings	No significant relation- ships appeared.
Comments: Sample grouped from 18 separate companies.					
Dicken & Black (1965)	Manufacturing Managers Insurance Managers	I. 31 II. 26	SVIB MMPI SVIB MMPI	Objective & Rating Criterion Objective & Rating Criterion	Some indication that physical science interests predicted negatively, but the relationships were not strong. Some indication of business interest predicting negatively.
Laurent (1962)	Managers--2 samples	200+ per sample	Guilford-Zimmerman Temperament Survey	Ratings	Highest r was +.23, with 5 other 19 r's .10 or more
Comments: Experimental keys only raised r to .24.					

11. Summary of Studies--Psychometric Prediction: Objective Personality and Interest Inventories (Cont.)

Investigators	Description of Sample and Study	N	Predictors	Criterion	Results
Robbins & King (1961)	Sales Managers--4 samples	100 to 538	a) KVPR b) Bernreuter Personality Inventory (BPI)	a) Position level changes b) Termination	a) Kuder predicted termination criteria significantly, but highest $r = .28$ b) Bernreuter showed similar results against termination criteria with the highest $r = .26$ c) Position level changes were not predicted by either criterion.
La Galpa (1960)	Naval officers--2 samples	Shore Sample, 299 Fleet Sample, 264	a) Cooperation Test b) Risk Scale	a) Ratings b) Critical Incidents Reports	Highest r was $+.30$, with no other r above $+.12$.

III. Summary of Studies--Psychometric Prediction: Personal History Data

Investigators	Description of Sample and Study	N	Predictors	Criterion	Results
Scollany (1957)	Promotion Managers	118	General PHB (Two Keys)	Ratings	Both Keys showed significant relationships.
Krauss (1964)	1st-line supervisors--4 samples	23 to 73	General PHB	Level changes	a) Correlations for PHB ranged up to the .50s. b) Age negatively related.
Meyer (1956)	1st-line supervisors	139	Age, Education	Ratings	Age positively related, Education non-related.
Comments: Possibility of criterion contamination.					
Mackinney & Wolins (1961)	Supervisors--3 samples	I. 66 II. 66 III. 45	Varied PHB items	Level, Tenure, Rankings & Suggestion Plan Activity	No consistent relationships found.
Thorndike & Hagen (1959)	Managers	1,055	General PHB	Self-described Pay, Mobility, Achievement	Results essentially negative.
Vernon (1950)	Civil-Service Managers (British)--2 samples	I. 147 II. 202	General PHB	Ratings	Results essentially negative.
Kirkpatrick (1966)	Chamber of Commerce Executives	161	General PHB	Salary	Significant relationships found.
Comments: Only extreme groups used.					

III. Summary of Studies--Psychometric Prediction: Personal History Data (Cont.)

Investigators	Description of Sample and Study	N	Predictors	Criterion	Results
Laurent (1962)	Managers--2 samples	200+ per sample	General PHB	Ratings	Highly significant correlations (.44 and .33)
La Gaipa (1960)	Naval officers--2 samples	Shore 229 Fleet 264	General PHB	a) Ratings b) Critical Incident Reports	Relationships were found to be insignificant.
Wollack & Guttman (1960)	Naval officers--2 samples	Fleet 108 Shore 81	Academic grades	Critical Incidents Reports	Relationships were found to be insignificant.
Ricciuti (1955)	Naval officers	308	Academic grades	Ratings	While several correlations were significant, none was above .19.
Kotula & Haggerty (1966)	Army officers--2 samples	426 in each group	General PHB	Ratings	Both correlations were significant, with highest absolute $r = .17$.
Haggerty (1953)	Army officers (Military Academy grads)	224	a) Academic grades b) Physical ed. grades	Combat rating criteria	a: highest correlation .13 b: two relevant correlations, .21 and .27
Haggerty (1963)	Army officers-- (Military Academy grads) 2 separate studies	I. 78 II. 420	a) Academic grades b) Physical ed. grades	a) Combat rating criteria b) Tenure	Highest correlation was .18

IV. Summary of Studies--Judgmental Prediction: Peer Ratings

Investigators	Description of Sample and Study	N	Criterion	Results
Ricciuti (1955)	Naval Officers	308 to 324	Overall Ratings	.32**, .22**, .33**
Williams & Leavitt (1947)	Marine Officers	100 100	Combat Ratings	.47** .43**
Hollander (1965)	Navy Officers--OCS Ratings	639	Overall Ratings	.24**, .40**, .39**
La Gaipa (1960)	Navy Officers-- Shore Duty Fleet Duty	10 10	Overall Ratings Overall Ratings Critical Incidents (Effective) Critical Incidents (Ineffective)	.26** .24** .29** -.16
Mollack & Guttman (1960)	Navy Officers-- Fleet Officers Shore Officers	108 81	Overall Ratings Critical Incidents (Effective) Critical Incidents (Ineffective) Critical Incidents (Average) Overall Ratings	.36** .15 .23* .23* .09
Haggerty (1953)	Military Academy Ratings	224	Overall Ratings	.28** to .33**
Haggerty (1963)	Military Academy Ratings Military Academy	78 420	Combat Criteria Rating Overall Ratings	.53** .33**

V. Summary of Studies--Judgmental Prediction: Superior and Faculty Ratings

Investigators	Description of Sample and Study	N	Predictors	Criterion	Results
Porter (1962)	Stanford MBAs	103	Faculty Ratings	Position level, Salary and several other objective and subjective criteria	Highest r was .26; 3 other r's were above .20.
Williams & Harrell (1964)	Stanford MBAs--2 samples 1) self-employed and family business and 2) employed by large org.	I. 196 II. 116	Faculty Ratings	a) Salary b) Administrative Level	1) Correlations with salary was .18*. 2) Correlation with salary was .03.
Williams & Leavitt (1947)	Marine Officers	100	Training Camp Leaders	Combat Ratings	1) Validity of ratings after 5 wks. was .22*. 2) Validity of ratings after 15 wks. was .36.
Ricciuti (1965)	Naval Officers	308 to 324	a) Naval Academy Officers b) Summer Cruise Officers	Ratings	r's ranged from .14 to .33.
Kotula & Haggerty (1966)	Army officers (ROTC)-- 2 separate studies	I. not given II. 426 in each of 2 samples	Faculty Ratings	Ratings	Highest correlation in either study was .23.
Haggerty (1953)	Army officers	224	Officer Ratings	Ratings	r's ranged from .21 to .26.
Haggerty (1963)	Army officers-- 2 separate studies	I. 78 II. 420	I. Faculty Ratings II. Officer Ratings	Ratings Ratings	I. r = .51 II. r = .37

9. Thomson, H. (1970). Comparison of predictor and criterion judgments of managerial performance using the multitrait-multimethod approach. Journal of Applied Psychology. 54(6), 496-502.

The multitrait-multimethod matrix technique was used to examine the predictive validity of ratings of management potential derived from an industrial assessment center program. Psychologists' and managers' ratings on 13 assessment dimensions were correlated with supervisors' ratings of current job performance on the same dimensions. Ratings obtained of on-the-job performance were lower in quality than the predictor ratings. The median reliability of the criterion ratings was .52 compared to median reliabilities of .85 and .89 for the psychologist's and managers' ratings, respectively. Furthermore, the supervisors failed to differentiate the various dimensions. The results of the present study were compared with findings of other studies that used this technique to determine the sources of unreliability in the criterion.

Type of assessment: Based on the Standard Oil model for the identification of managerial potential. A 3 day program with 2 staff psychologists, and 3 trained managers serving as observers-assessors

Tests: objective tests (paper and pencil) and projective tests (undesigned), simulated management exercises, interviews, and an oral presentation

Predictive validity: The summary rating of the assessors on 13 dimensions was designated to predict future on-the-job behavior, or managerial potential. Criterion: supervisors' ratings on the same 13 dimensions; six months to 2 1/4 years later)

Management level: Professional, technical, and lower level management (The average age was 32 years; most were college graduates, six years plus service with the company)

The two assessment methods (psychologists vs. managers as assessors) showed very high concurrent validity ($r_{mn}=.85$), while the predictive validities of both psychologists' and managers' ratings with the supervisors' ratings were substantially lower (.42 and .38). Nevertheless, for both predictor methods, the criterion-related validity coefficients were significant for 11 of the 13 traits.

Thomson (continued)

Validation of Assessment Ratings

<u>Trait</u>	<u>Psychologists - Managers</u>	<u>Psychologists - Supervisors</u>	<u>Managers - Supervisors</u>
Amount of Participation	.92	.58	.65
Oral Communication	.82	.44	.37
Personal Acceptability	.78	.33	.37
Impact	.88	.50	.47
Quality of Participation	.82	.31	.38
Personal Breadth	.86	.55	.52
Orientation to Detail	.79	.19	.14
Self-Direction	.89	.42	.35
Relationship with Authority	.85	.35	.37
Originality	.87	.45	.45
Understanding of People	.79	.35	.39
Drive	.73	.12	.29
Potential	.93	.64	.64
Median	.85	.42	.38

Analysis for convergent and discriminate validities reveals that the assessment center raters were able to discriminate among the various dimensions despite the relatively high level of intercorrelations between traits. However, the supervisors' ratings had failed to adequately discriminate among the different assessment dimensions.

The high concurrent validity of the psychologists' and managers' assessment ratings could have been attributable to the extensive training given these two groups in the assessment process. In contrast, the criterion raters received less training on the meaning and use of the scales and judged a fewer number of subjects within the context of differing job requirements. These factors apparently made it difficult for the criterion judges (supervisors) to establish a common frame of reference for making the ratings.

10. McConnell, J. & Parker, T. (1972). An assessment center program for multi-organizational use. Training and Development Journal. 26(3), 6-14.

This program was modelled after the Wolverine Tube Division of Universal Oil Products Company (1967) and was designed to simplify the training of assessors and the administration of the program. The results were reported to be equal to or superior to the published results of larger and more complex programs. Using this experience, the rapid innovations in multi-media training, and a proven method of providing organizations with materials and know-how for successfully conducting in-house programs, a multimedia assessment center program designed to be adaptable to most organizations was designed and tested by DKD Inc. for the American Management Association.

Type of assessment: identification of first level management ability and potential and evaluation for developmental purposes. A one-day series of exercises (Management Simulation workshop):

McConnell and Parker (continued)

Tests: 1) An interview (based on background information), 2) Management Questionnaire, 3) In-Basket, 4) Luncheon meeting, 5) Film Case Discussion (employee promotion), 6) Selection Simulation, 7) Management Decision Game, 8) Participant's Written Evaluations of Performance Level: first level supervisor

Validity and reliability:

- I. Reliability Results: (how well do the assessors agree on judgments)
 1) Internal consistency reliability estimates within assessment categories:

(N = 5 assessors for each 12 participants)

<u>Assessment Category</u>	<u>Reliability Coefficient</u>
Functional Ability	.85
Planning Ability	.78
Organizing Ability	.83
Controlling Ability	.76
Oral Communication Ability	.79
Written Communication Ability	.90
Leadership Ability	.90
Company Orientation	.86
Decision-making Ability	.77
Creative Ability	.64
Initiative	.83
Flexibility	.76
Overall Management Ability	.89
Management Potential	.83

- 2) Internal Consistency Reliability Estimates of Overall Management Ability Within Groups (N=5 assessors)

<u>Field Test Group</u>	<u>N Participants</u>	<u>Reliability Coefficient</u>
Bank	33	.94
Paper Manufacturer	12	.98
Electronics Manufacturer	48	.91
Auto Manufacturer	12	.85
Textile Manufacturer	12	.90
Retail Store	12	.89
Total	129	.90

McConnell and Parker (continued)

3) Estimates of Test-Retest Reliability for Overall Management Ability in Two Field Test Situations:

<u>Field Test Group</u>	<u>N Participants</u>	<u>Test-Retest Reliability</u>
Bank	9	.79
Electronics Manufacturer	12	.73
Total	21	.74

(Used the same groups of participants on two separate occasions with two independent groups of assessors.)

II. Validity Results:

1) Concurrent Validity Coefficients for the Overall Management Score from the Assessment Procedure: (criteria: performance ratings independently collected)

<u>Field Test Situation</u>	<u>N Participants</u>	<u>Correlation Coefficient</u>
Bank	24	.55
Paper Manufacturer	12	.48
Electronics Manufacturer	22	.64
Auto Manufacturer	12	.28
Total	70	.57

2) Accuracy of Assessment Program in Identifying Job Performance Ratings:

<u>Field Test Situation</u>	<u>N Participants</u>	<u>Correct Identifications</u>	<u>Misses +/-1 Scale Pt.</u>	<u>Misses +/-2 Scale Pts.</u>
Bank	24	50%	42%	8%
Paper Manufacturer	12	50%	42%	8%
Electronics Manufacturer	22	64%	32%	4%
Auto Manufacturer	12	75%	24%	—
Total	70	59%	36%	5%

11. Kraut, A. & Scott, G. (1972). Validity of an operational management assessment program. Journal of Applied Psychology. 56(2), 124-129.

The validity of an assessment program is examined by reviewing the career progress of 1086 employees in sales, service and administrative functions after they were assessed. Although participants were nominated on the basis of being promotable, raters found that more than one-quarter were unqualified, and the others were widely differentiated. Ratings were used to move men into first-line management, but the relationship of ratings to first promotions is moderate enough to reduce fears of "crown prince" or "kiss of death" effects. Nor does participation seem to demotivate these employees. Assessment ratings are substantially correlated with two major criteria, second-level promotions and demotions from first-line management.

Type of assessment: Traditional techniques include 2 1/2 day program, plus two days of developmental activities; for assessing management potential

Tests: Paper and pencil tests (undesignated, but on AT&T model)

Situational exercises, including leaderless discussion group and in-basket exercise

Developmental exercises: lectures, seminars and classroom discussions

Feedback interview

Management level: pre-management level

Predictive validity: (N=1086) Based on differentiation of management potential according to assessment ratings, second promotions were proportionately greater for those rating higher in the assessment program; demotions were proportionately larger for those rated lowest.

12. Byham, W. (1980). The assessment center as an aid in management development. Training and Development Journal. 34(6), 24-36.

A reprint of his earlier classic article with a 1980 update. Contents: a typical center, early identification, individual management development needs, stimulation of self-development, assessor training, assessor level, validity and relation to EEO guidelines, pure research studies, operational studies, IBM studies, negative studies, differential validity, problems in assessment centers, anxiety, costs, update.

Type of assessment: Traditional or classic approach

Management level: From entry level to four levels above entry management

Predictive validity: The article cites the AT&T Management Progress Study - 20 year follow-up as indicative that assessment center can validly predict management success as high as four levels above entry management positions; overall rating .51, paper and pencil tests .33.

13. Cohen, S. (1980). Pre-packaged vs. tailor-made: The assessment center debate. Personnel Journal. 59(12), 989-991.

Four major tested assessment techniques:

- 1) Analysis of the (observed) exhibited behavior of an individual
- 2) Development of simulated exercises representing work activities
- 3) Determination of relative skill strengths and weaknesses
- 4) Transmission of objective results of assessment to individual

Basic steps in setting up an assessment center:

- 1) Job analyses
- 2) Skill identification
- 3) Exercise design
- 4) Assessor training
- 5) Center administration/delivery
- 6) Observation and evaluation (only on job relevant and readily observables and by multiple assessors) (prepackaged programs generally means using a behavior checklist rather than open ended observation)
- 7) Feedback
- 8) Validation

Assessor training: Prospective assessors should participate in the exercises and agree to the standards of performance that are to be used to evaluate candidates.

Type of assessment: pre-packaged vs. tailor-made

Unwanted differences characteristic of pre-packaged programs:

- 1) Targeted job analysis not done
- 2) Skills to be identified and evaluated are dictated by the program
- 3) Exercises - may not be realistic simulations of the specific environment
- 4) Assessors - may not be sensitive to the unique requirements of the organization
- 5) The program, administration and delivery may not coordinate or collaborate with established company procedures and practices
- 6) Evaluation procedures are standardized, may be less sensitive than less restrictive, open ended evaluation
- 7) Feedback may not be relevant if exercises have not been relevant
- 8) Requires minimum in-house assessment expertise

Management level: not designated

Validity: The author remarks that validity would depend on the relevance of the exercises to the organizational situation.

14. Souder, W. & Leksich, A. (1983). Assessment centers are evolving toward a bright future. Personnel Administrator. 28(11), 80-87.

The authors discuss the past, present and future of assessment centers under the topics: modern form, exercise selection, assessor training, feedback procedures, validation, systems design, current trends.

Type of assessment: Traditional and/or classic

Tests: Multiple methods that simulate real world environments:

objective tests, projective tests, interviewing, face-to-face exercises; problem solving situations; peer evaluations and trained, professional evaluations; feedback

Variations on classic model of tailor-made, in-house center:

- 1) Contractor-operated: pre-packaged programs with standard videotapes, cassettes, tests, exercises, and evaluation forms
- 2) Community center - a jointly owned or leased facility, shared by several organizations; may be customized for a particular organization, or standardized.
- 3) Second generation assessment centers: highly automated variations of the pre-packaged center - i.e., participants follow videotaped instructions and dialogue with a computer - the effect of the absence of real human contact has not been determined.

Management level: Not designated, presumably all levels

Predictive validity: The critical elements for a center to be successful were identified:

- 1) A system perspective integral to specific system and consistent with its policies, goals and culture
- 2) Top management support and involvement, providing objectives, criteria and feedback
- 3) Careful selection of candidates according to pre-specified objectives and criteria
- 4) Maintain valid exercises, competent assessors, and honest feedback to maximize credibility
- 5) Assessment center viewed and managed as an agency of change

APPENDIX B

The following studies are reviewed in Appendix B. They reflect validity data of specific assessment exercises. They are presented in chronological order.

1. Frederickson, N., Saunders, D. & Wand, B. (1957). The in-basket test. Psychological Monographs. 71(9), 1-28.
2. Ginsburg, L. & Silverman, A. (1972). The leaders of tomorrow: Their identification and development. Personnel Journal. 51(9), 662-666.
3. Ulrich, L. & Trumbo, D. (1965). The selection interview since 1949. Psychological Bulletin. 63(2), 100-116.
4. Bray, D. & Grant, D. (1966). The assessment center in the measurement of potential for business management. Psychological Monographs. 80(17), 1-27.
5. Grant, D., Katkovsky, W. & Bray, D. (1967). Contributions of projective techniques to assessment of management potential. Journal of Applied Psychology. 51(3), 226-232.
6. Greenwood, J. & McNamara, W. (1967). Interrater reliability in situational tests. Journal of Applied Psychology. 51(2), 101-106.
7. Hardesty, D. & Jones, W. (1968). Characteristics of judged high potential management personnel: The operations of an industrial assessment center. Personnel Psychology. 21(1), 85-98.
8. Grant, D. & Bray, D. (1969). Contributions of the interview to assessment of management potential. Journal of Applied Psychology. 53(1), 24-34.
9. Hinrichs, J. (1969). Comparison of 'real life' assessments of managerial potential with situational exercises, paper-and-pencil ability tests, and personality inventories. Journal of Applied Psychology. 53(5), 425-432.
10. Wollowick, H. & McNamara, W. (1969). Relationship of the components of an assessment center to management success. Journal of Applied Psychology. 53(5), 348-352.
11. Carlton, F. (1970). Relationships between follow-up evaluations and information developed in a management assessment center. Proceedings, 78th Annual Convention, American Psychological Association, 565-566.
12. Dodd, W. (1970). Will management assessment centers issue selection of the same old types? Proceedings, 78th Annual Convention, APA, 569-570.
13. Finley, R. Jr. (1970). Evaluation of behavior predictions from projective tests given in a management assessment center. Proceedings, 78th Annual Convention, APA, 567-568.

APPENDIX B

Validity Studies of Specific Assessment Exercises

1. Frederickson, N., Saunders, D. & Wand, B. (1957). The in-basket test. Psychological Monographs. 71(9), 1-28.

The article is a complete and detailed review of the purpose, description, development of the in-basket test, containing the following sections:

Introduction

General Description of the In-Basket Test

How the In-Basket Test was Developed - Use of Essays, Interviews in constructing problems, steps in problem preparation, development of administrative procedure, scoring procedure

Results of the Command and Staff School Tryout - scoring reliability, all problems ($r = .90$), overall reliability (too complex to report), validity (no conclusive evidence can be offered). In the final analysis, validity is confidence in a test which is generally borne out...over a period of time, i.e., prevailing attitudes towards the test

Recommendations and suggestions for further research, including recommendations for use of the test in its present form are provided together with a Summary and Appendix with sample problems.

2. Ginsburg, L. & Silverman, A. (1972). The leaders of tomorrow: Their identification and development. Personnel Journal. 51(9), 662-666.

In developing a program to ensure the building of solid first and second level management teams, several parameters were established: the program must validly measure management potential; have high acceptability by all concerned; be administered as an integral part of the organization's management development program; be flexible and comprehensive; and have a high payoff value. The result was what may be the first hospital personnel identification and development center.

Type of assessment: Identification of strengths and developmental needs of administrative personnel, one day program

Tests: paper and pencil tests (not specified), patterned interview (semi-structured), situational tests and exercises: in-basket, leaderless group discussion

Management level: First and second level management (N=54)

Predictive validity: (Criterion: ratings of performance by supervisors - an overall leadership evaluation)

Best predictors of general leadership performance:

analytical thinking: .34
decision making (In-Basket): .37

Best predictors of planning and organizing skill:

decision making (In-Basket): .35
analytical thinking: .34
Interview impressions of planning and organizing: -.34 (surprising)

Best predictor for ability and desire to communicate with subordinates and supervisors:

ability to think logically: .34

Best predictor in human relations area:

(In-Basket) human relations: .31

3. Ulrich, L. & Trumbo, D. (1965). The selection interview since 1949. Psychological Bulletin. 63(2), 100-116.

Research literature on the selection interview since 1949 is reviewed. Major sections include validity studies, studies dealing with the accuracy of information obtained in the interview, and analytic and model-testing studies. Validity evidence is generally confounded in that the predictions which are validated are made on the basis of both face-to-face interview and other ancillary data. However, the bulk of the evidence favors both the structured interview and interviews limited in purpose. Recurring evidence suggests that the interview may be most successful if limited to the assessment of personal relations and career motivation. Recent analytic studies involving content analyses and decision-making processes show promise of providing new insights into the interview process.

4. Bray, D. & Grant, D. (1966). The assessment center in the measurement of potential for business management. Psychological Monographs. 80(17), 1-27.

Contains a full report of the Bell Systems Management Progress Study, description and results of analyses. The article contains:

Nature of assessment, experience with assessment, theoretical and methodological considerations

Assessment procedures, techniques, administration and reporting, rating variables, staff evaluations

Analyses of Staff Evaluations, method, results

Analyses of Techniques, group exercises, rater agreement, method agreement, overlap of exercises, oral communications skills, contributions to staff evaluations, in-basket, mental abilities tests, personality and attitude questionnaires, relative contributions

Prediction of Progress, specific variables, assessment techniques

Discussion - nature of the judgments, contributions of the techniques, validity of evaluations

Bray and Grant (continued)

Type of assessment: Classic; for potential success in business management or for developmental potential; a 3 1/2 day program

Paper and pencil tests: SCAT, form I; Test of Critical Thinking; Contemporary Affairs Test, Edwards Personal Preference (EPPS), Guilford Martin Inventory (GAMIN), Opinion Questionnaire (Bass), Attitude Survey

Situational exercises: In-Basket Manufacturing Problem, Group Discussion

Projectives: Rotter Incomplete Sentences Blank, Bell Incomplete Sentences Test, 6 cards of TAT
Two hour interview
Other: personal history questionnaire, autobiographical essay; Q sort - self-descriptive

Level of management: Entry and lower level management, predicting to middle management

Predictive validity: See tables below

Conclusions: The data presented appear to justify the costs entailed.

The data reported make it apparent that the situational techniques (group exercises and In-Basket) used, despite their complexities, produced reasonably reliable results and that they markedly influenced the judgments of the assessment staffs. The paper and pencil instruments had less influence on staff evaluations generally, though they did influence them in many specific ways.

The findings further indicate that neither kind of technique could have been omitted without loss of important information.

Progress in Management

<u>Educational Background</u>	<u>Percentage at Each Management Level</u>			
	<u>N</u>	<u>3-4</u>	<u>2</u>	<u>1</u>
College	125	30	64	6
Non-College	144	13	42	45
College and Non-College Combined	269	21	52	27

Levels 3 and 4 are middle management levels. Level 3 is the level which college graduate management trainees are expected to achieve within 5-10 years.

Bray and Grant (continued)

Relationship of Staff Predictions to Progress

<u>Sample</u>	<u>Staff Prediction</u> (will attain middle mgmt)		<u>Achieved Mgmt Level (%)</u>			<u>Significance</u>
			<u>3-4</u>	<u>2</u>	<u>1</u>	
College	yes	62	48	50	2	.001
	no or?	63	11	78	11	
Non-College	yes	41	32	61	7	.001
	no or?	103	5	35	60	
Total Combined	yes	103	42	54	4	.001
	no or?	166	7	51	42	

*used χ^2 test or "Fisher's exact test"

Correlations with Salary Progress

Predictor variable	SAMPLE						
	A (N = 54) (College)	B (N = 83) (Non-College)	C ₁ (N = 27) (College)	C ₂ (N = 39) (Non-College)	D ₁ (N = 19) (College)	D ₂ (N = 22) (Non-College)	E (N = 25) (College)
Staff judgment							
General effectiveness (I)	41*	45*	51*	52*	24	13	34
Administrative skills	33*	57*	24	45*	32	-11	24
Interpersonal skills	26	34*	36	33*	29	28	40*
Control of feelings	34*	17	50*	32*	20	04	00
Intellectual ability	48*	31*	30	07	30	-13	18
Work-oriented motivation	16	29*	20	41*	05	35	15
Passivity	-30*	-41*	-33	-41*	21	-15	-40*
Dependency	-25	-01	-25	01	07	24	07
Nonconformity	34*	--	32	--	16	--	20
Situational exercises							
Manufacturing Problem	15	37*	41*	50*	14	29	-01
Group Discussion	30*	33*	50*	28	26	10	38
In-Basket	27*	44*	-01	22	03	-19	28
Ability Test							
SCAT verbal	36*	35*	51*	30	19	-44*	14
SCAT quantitative	23	44*	-04	19	09	-10	-28
SCAT total	38*	45*	32	28	18	-30	-03
Critical thinking in social science	26	46*	-21	36*	-02	-38	29
Contemporary affairs	35*	26*	32	-09	32	-17	37
Questionnaire							
Edwards PPS							
ach	20	09	12	25	-15	-28	-10
def	-03	09	13	-19	-06	02	42*
ord	-05	-01	-09	21	15	39	-23
exh	-03	-03	20	25	-38	-21	17
aut	-09	-01	-04	-07	01	-55*	04
aff	-12	-11	-01	00	-25	16	18

Correlations with Salary Progress

Predictor variable	SAMPLE						
	A (N = 54) (College)	B (N = 83) (Non-College)	C ₁ (N = 27) (College)	C ₂ (N = 39) (Non-College)	D ₁ (N = 19) (College)	D ₂ (N = 22) (Non-College)	E (N = 25) (College)
Questionnaire (cont.)							
Edwards PPS							
int	14	24*	05	14	02	35	12
suc	-14	-29*	-15	-16	08	09	-19
dom	26	40*	01	-05	10	19	27
aba	-32*	-25*	13	11	-08	-11	07
nur	-02	-28*	02	-14	-23	30	-01
chg	01	10	-41*	-11	13	-22	03
end	02	-07	-39*	03	15	39	00
het	01	-05	13	-15	13	-01	-39
agg	15	17	49*	01	13	-43	-03
Gilford-Martin							
G	-03	--	24	02	32	25	22
A	20	--	26	12	35	22	09
M	17	--	-45*	-05	14	13	06
I	19	--	-03	08	-10	16	-03
N	-07	--	-02	18	-36	22	05
Attitudes toward life	08	05	52*	-06	08	23	-01
Opinion questionnaire							
A	02	-24*	10	-15	-22	12	-25
a	-18	-02	-07	-27	-19	31	-14
n	08	-06	05	09	-01	-25	-42*

*P less than .05 than $r = .00$.

NOTE: D₁, D₂, and E comprised of people who had at least six years of management experience.

Partial Correlations with Salary Progress

	<u>Sample Groups</u>			
	<u>College (N = 54)</u>	<u>Non-College (N = 83)</u>	<u>College (N = 27)</u>	<u>Non-College (N = 39)</u>
	<u>Partial r</u>	<u>r</u>	<u>Partial r</u>	<u>Partial r</u>
Staff judgment	.48	.57	.51	.52
	.32*	.39**	.29	.42**
Ability test	.38	.46	.51	.36

*p .05

**p .01

Partial r - when mental ability, measured by paper-pencil test, is partialled out of judged ability. The remaining reliable variance indicates that the assessment process contributes more than can be gained by simple paper and pencil ability measures.

5. Grant, D., Katkovsky, W. & Bray, D. (1967). Contributions of projective techniques to assessment of management potential. Journal of Applied Psychology. 51(3), 226-232.

The contributions of projective techniques to assessment-center staff evaluations and the relationships of projective variables to progress in management are presented. The projective data were obtained by coding reports written by a clinical psychologist from three projective instruments. Analysis of the data show that the projective reports particularly influenced the assessment staff in rating such characteristics as work motivation, passivity, and dependency. In addition, several of the projective variables are reliably related to progress in management, especially those pertaining to leadership and achievement motivation. In brief, the findings clearly indicate that relevant information on managerial motivation was obtained from the projective reports.

Type of assessment: For prediction of management potential; traditional assessment techniques (projective techniques were examined in this analysis)

Tests: Paper and pencil tests: (ATT Management Progress Study program tests)
Interview (only referred to indirectly)
Projective tests: Rotter Incomplete Sentences Blank; Management Incomplete Sentences Test; 6 cards of TAT (cards 2, 6BM, 7BM, 8BM, 14 and 16)

Management level: Early management or pre-management

Predictive validity: (N=207 and N=148) The contributions (correlations) of various assessment techniques to the staff judgment of overall prediction. (For the college educated group only):

Manufacturing Problem	.41	Personality Questionnaire	.29
Group Discussion	.60	Projective Report	.35
In-Basket Exercise	.55		
Mental Ability Test	.36		

Predictive validity: The correlations of salary progress with elements of the projective report - (college sample only)

Leadership role	.24
Dependence	-.35
Subordinate role	-.25
Achievement motivation	.26

6. Greenwood, J. & McNamara, W. (1967). Interrater reliability in situational tests. Journal of Applied Psychology. 51(2), 101-106.

This study was conducted to determine the degree of interrater reliability in situational tests and to determine the relative effectiveness of professional and non-professional evaluators in this type of situation. The results indicate that the reliability of observer ratings and rankings are reasonably high in several different situational tests. Of particular significance is the finding that adequate reliability can be obtained from the use of non-professional evaluators in business-oriented situational tests.

Type of assessment: Managerial selection - two day program;

Tests: Paper and pencil tests: traditional, but undesignated

Situational tests: Leaderless Discussion Group; Task Force Exercise; manufacturing exercise

Other: undesignated, but a "traditional" personnel assessment program

Management level: Lower and middle management; and middle to upper management

Predictive validity: Statistical evaluation only for inter-rater reliability on situational tests: (N=288), range of coefficients: .48 to .89, most in the .70s to .80

Overall evaluation: observer ratings and rankings have reasonable reliabilities in most cases.

7. Hardesty, D. & Jones, W. (1968). Characteristics of judged high potential management personnel: The operations of an industrial assessment center. Personnel Psychology. 21(1), 85-98.

The contents of this article cover the industrial assessment center, an analysis of one program, and additional comments and observations. Results: judged high potential individuals rated significantly higher on scales of business motivation, oral communication assertiveness, and compatibility. They were younger, better educated, more athletic (as a high school activity), held more student government positions, more were "A" students, changed residence moderately rather than much or not at all, had been military officers, had professional or managerial fathers, had more working mothers, had fathers with college degrees and/or graduate courses, tended to rise faster in their organization than peers, had better company records. However, there were many individuals who were considered "finds."

Type of assessment: Independent evaluation of potential and developmental needs (three day, live-in program)

Tests: Paper and pencil tests: SCAT; Doppelt Math Reasoning Test, Minnesota Engineering or Millers Analogies Test; Davis Reading Comprehension and Speed Tests, Watson-Glaser Critical Thinking, Dogmatism Test, Gordon Interpersonal Values, Michigan Vocabulary Test, Interest Checklist, and background information

Situational exercises: Stock Market Exercise and two others (unnamed)

Projective tests: TAT and Sentence Completion

Interview: (conducted by three top level managers in the (assessment) program) and four sociometric rankings were assigned.

Management level: Young professionals and managers (ages 25-40); selected as high performers

Predictive validity: Based on relative frequency differences between High Potential and Not High Potential managers. Assessment was judged a valid process as it represented a reasonable improvement over traditional approaches of information gathering and decision making.

8. Grant, D. & Bray, D. (1969). Contributions of the interview to assessment of management potential. Journal of Applied Psychology. 53(1), 24-34.

The contribution of interview information to assessment center evaluations and the relationship of interview variables to progress in management are presented. The interview data were obtained by coding interview reports. Analysis of the data clearly indicate that information from the interview reports contributes to assessment center evaluations. Judgments of career motivation and, to a lesser extent, work motivation and control of feelings appear to have been influenced by the interview information. In addition, judgments of interpersonal skills were reinforced, if not influenced, by the interview reports. The results of the study also demonstrate that extensive and reliable information on many personal characteristics can be obtained from the interview. In addition, several of the interview variables, especially those reflecting career motivation, dependency needs, work motivation, and interpersonal skills are directly related to progress in management. The findings clearly indicate that relevant information on personal characteristics important to managerial success was obtained from interview reports.

Type of assessment: Traditional, for identification of management potential (The focus of the study was on the contribution of the interview.)
Traditional assessment techniques, including a semi-structured 2-hour interview by professional psychologists

Management level: Generally early management level

Predictive validity: (N=200 (college graduates) and 148 (non-college))
Criterion: Reaching middle management within 10 years; Predictors: interview variables

	<u>Correlations with Staff Predictions</u>		<u>with Salary Progress</u>	
	<u>College Graduate</u>	<u>Non-College</u>	<u>College Graduate</u>	<u>Non-College</u>
Personal Impact (Form)	.49*	.21	.28*	.17
Oral Communication Skills	.41*	.48*	.22	.50*
Human Relations Skills	.23*	.38*	.20	.41*
Personal Impact (Likable)	.25*	.14	.22	-.11
Behavior Flexibility	.19*	.11	.30*	.04
Need Approval - Superiors	-.02	-.20*	-.36*	-.27*
Need Approval - Peers	-.21*	-.13	-.36*	-.17
Tolerance of Uncertainty	.23*	.36*	.06	.23*
Inner Work Standards	.17*	.07	.07	.08
Primacy of Work	.20*	.21	.30*	.25*
Energy	.25*	.17	.35*	.16
Goal Flexibility	-.30*	-.21*	-.08	-.16
Need Advancement	.28*	.42*	.49*	.44*
Need Security	-.28*	-.17	-.35*	-.26*
Social Objectivity	.03	.18*	.17	.20*
Company Value Orientation	-.13	-.05	-.11	-.10
Ability to Delay Gratification	.01	.03	-.16	.12
Range of Interest	.27*	.26*	.28*	.22*

*p = <.05

9. Hinrichs, J. (1969). Comparison of 'real life' assessments of managerial potential with situational exercises, paper-and-pencil ability tests, and personality inventories. Journal of Applied Psychology. 53(5), 425-432.

The major components of evaluation from a two-day assessment program covering 47 members of a large national marketing organization consisted of ratings of the degree of active participation in the group situational exercises, followed by ratings of administrative and decision-making ability. Paper-and-pencil ability tests and personality inventories were less clearly related to assessments of managerial-potential. Ratings of management potential developed from a careful review of company personnel records were as highly correlated with the assessment center data as were overall ratings from the two-day program, except for ratings dealing with interpersonal behavior.

Type of assessment: Similar to (Bray and Grant's) traditional program, two days, for assessing promotion potential

Tests: Paper and Pencil tests: Concept Mastery Test, Form T; SCAT, Form V; Gordon Personal Profile; Allport, Vernon and Lindzey Study of Values; Leadership Opinion Questionnaire, Ghiselli Self Description Inventory; Risk-Taking Scale; background information
Situational Exercises: Leaderless Discussion Group, Task Force Committee; Manufacturing Game; In-Basket; Stock-Market (individual) and job environment report (individual)

Management level: Pre- and lower-level management

Predictive validity: (N=47) 1) Concurrent, not predictive, with evaluations based on personnel folder data and current salary standing (for general management potential): .46 (organizational data); .37 (for salary standing), 2) Contribution (correlation) of overall assessment rating with contributing rating scales: activity .78; administration .50; stress resistance .25, 3) Added value of assessment technique to evaluation of interpersonal relationships - the contribution (correlation) of interpersonal relationship factors to the overall evaluation:
- by assessment center technique .78
- by evaluation of personnel folder .49

Construct validity: overlapping of variables measured by separate tests, i.e., situational vs. personality tests
For - interpersonal activity: ascendancy .56; political values .38; self-confidence .43;
For - stress resistance: occupational level .44; initiative .36; risk-taking .36

Conclusion: "The major unique component of assessment, both from the situational and from the personality tests, seems to be an evaluation of interpersonal behavior."

10. Wollowick, H. & McNamara, W. (1969). Relationship of the components of an assessment center to management success. Journal of Applied Psychology. 53(5), 348-352.

This study determined the validity of an assessment center approach in predicting management potential and the relative value of the components of the program. Results indicate that the approach is valid and that situational tests add to predictiveness through statistical combination of the program variables, rather than a subjectively derived overall rating.

Type of assessment: Early identification of management potential; all designated as having above average potential for advancement - two-day, "traditional" program.

Tests: Paper and Pencil Tests: Gordon Personal Profile and Inventory, Fleishman's Leadership Questionnaire; SCAT: Otis Employment Test; Background and Contemporary Data Form
Situational Tests: Leaderless Group Discussion, Manufacturing Exercise, Task Force Exercise; In-Basket exercise; job environment report (individual); stock market exercise (individual)

Management level: Lower and middle managers with above average potential

Wollowick and McNamara (cont.)

Predictive validity: For increased managerial responsibility approximately three years later (N=94)

Paper and pencil tests	.45	Tests and Exercises	.54
Exercises, situational	.39	Characteristics and Exercises	.52
Judged Characteristics	.41	Characteristics and Tests	.55
Tests, characteristics and exercises			.62

Predictive validity: Specific contributions (correlations) with the success criterion (later increased management responsibility):

Tests: Ascendancy	.39
Sociability	.23
Vigor	.32

Exercises: Manufacturing	.28
Leaderless Disc.	.25
In-Basket	.32

Background: Self-confidence	.23
-----------------------------	-----

Predictive validity: Specific contributions (correlations) with the success criterion (later increased management responsibility): (cont.)

Personal characteristics based on the exercises:

Self-confidence	.32
Written Communication	.29
Energy Level	.26
Decision Making	.29
Resistance to Stress	.26
Planning/Organizing	.23
Persuasiveness	.22
Aggressiveness	.24
Oral Communications	.22

Overall rating score for total assessment: .37

11. Carlton, F. (1970). Relationships between follow-up evaluations and information developed in a management assessment center. Proceedings, 78th Annual Convention, American Psychological Association, 565-566.

Type of assessment: Standard Oil Company - Formal Assessment of Corporate Talents (FACTs) Program - planned follow-up research, 2 1/2 to 5 years after assessment

Standardized rating of characteristics:

- | | |
|-----------------------------|------------------------------|
| 1) amount of participation | 8) self-direction |
| 2) oral communication | 9) relationship to authority |
| 3) personal acceptability | 10) originality |
| 4) impact | 11) understanding people |
| 5) quality of participation | 12) business drive |
| 6) personal breadth | 13) potential |
| 7) orientation to detail | |

Criterion: 1) Job performance on the 13 characteristics judged by two key managers, 2) composite criterion - an index based on salary growth and promotions over a four year period

Predictors: Committee rating, projective rating (TAT & Sentence Completion), and interview rating

Other variables: Best scores and personal history variables

Generally, the follow-up ratings correlated significantly and moderately well (.30s and .40s) with the composite criterion, suggesting that the characteristics are related to managerial performance (see table for details).

Type of assessment: Traditional, management potential

Tests: Paper and pencil: Millers Analogy, Michigan Vocabulary, General Information Test, SCAT - verbal and quantitative, Watson-Glaser Critical Thinking, Doppelt-Mathematics, Davis Reading - speed and comprehension, A.C. Creative Test (Quantitative), Welsh Figure Preference, Gordon Personal Profile
Projective techniques - not specified
Interview
Background Information

Management level: Early; predict to job performance and success criterion 2 1/2 to 5 years later

Predictive validity: See table below

Correlations between Assessment Center Variables and Follow-Up Manager's Ratings

Variable	N	Comp. Crit.	Participation, Amount	Communication, Oral	Acceptability, Personal	Quality of Participation	Orientation to Detail	Self-direction	Relationship to Authority	Originality	Understanding People	Business Drive	Potential
Composite Criterion	122		.33*	.30*	.13	.44*	.32*	.19*	.44*	.42*	.26*	.40*	.63*
Full Committee	122		.35*	.26*	.31*	.24*	.29*	.01*	.24*	.28*	.20*	.29*	.65*
Projective report	122		.24*	.11	.13	.19*	.25*	.11	.14	.18*	.13	.12	NA
Interview report	122		.28*	.05	.16	.25*	.25*	.16	.31*	.36*	.22*	.33*	NA
Salary grade	122		.19*	.19*		.26*	.26*	.20*	.23*	.27*	.31*		.37*
Education	122	.22*				.19*							.29*
Tenure	122	-.36*			-.19*					-.21*			-.19*
Millers Analogy	122	.37*		.20*		.33*			.26*	.32*			-.31*
Michigan Vocabulary	76	.34*	.23*	.22		.40*	.38*		.29*	.37*	.23*		.40*
General Information Test	52			.23		.21	.20		.21	.29*	.25		.27*
SCAT - Verbal	107	.29*		.30*		.35*			.29*	.30*			.29*
Watson-Glaser	122	.25*		.21*		.23*			.21*	.20*			.23*
Doppelt-Mathematics	122	.29*				.28*			.18*	.25*			.30*
SCAT - Quantitative	107	.36*		.19*		.32*	.18*		.26*	.28*			.33*
Davis Reading - Speed	122	.28*				.25*			.24*	.25*			.21*
Davis Reading - Level	122	.28*				.28*			.25*	.22*			.23*
A.C. Creative (Quantitative)	45		.34*						.22				
Welsh Figure	21	.43*		-.40		.43*			.36	.41			.20
Preference (Art)													
Welsh Figure Preference (Conform)	21	-.72*							-.33	.37	-.26	-.27	-.48*
Gordon - Conform	105	-.26*				-.39	-.25*						
Gordon - Benevolence	105										.20		
Gordon - Leadership	105	.23*											

*p / = .05

12. Dodd, W. (1970). Will management assessment centers issue selection of the same old types? Proceedings, 78th Annual Convention, APA, 569-570.

This study examines validating early management identification procedures against supervisory recognition of management talent and performance in an assessment center. To the extent that validations against the two criteria predict different types of people to be successful, the assessment center represents a change away from usual standards of supervisory recognition.

Type of assessment: Management potential; 2 1/2 day AT&T model

Management level: Select first line managers (N=573)

Predictive validity: Criteria: 1) Attendance at assessment center (supervisory recognition) vs. unassessed managers, 2) Management potential ratings of assessment attendees.

Discussion: With the promotion system as it stands, choosing to seek out additional management candidates by an early identification program would result in different types of candidates for promotion if the criteria were supervisory recognition rather than if the criterion were assessment performance after such recognition. Of those predicted to be top assessment performers (in the top two ratings in advanced training and having high scores on ascendancy), not all had been selected for the program. These groups were not as low on independence and were much lower on conformity than those recommended to be assessed.

Relationship of Early Identification Measures to Later Criteria
of Management Potential

Early Test and Training Measures	Differences (in Means) Between Assessed Salesmen and Unassessed Surviving Salesmen	Correlation with Overall Assessment Rating
Survey of Interpersonal Values		
Support	-1.3	- .20
Conformity	1.0	- .17
Recognition	.5	- .02
Independence	-2.4*	.15
Benevolence	- .3	.19
Leadership	2.1*	.09

Dodd (continued)

<u>Early Test and Training Measures</u>	<u>Differences (in Means) Between Assessed Salesmen and Unassessed Surviving Salesmen</u>	<u>Correlation with Overall Assessment Rating</u>
Gordon Personal Inventory		
Caution	- .1	- .14
Original Thinking	.5	.12
Personal Relations	- .2	.07
Vigor	1.3	.10
Gordon Personal Profile		
Ascendancy	.9	.27*
Responsibility	.4	- .17
Emotional Stability	- .7	- .11
Sociability	.8	.25
Basic Sales Training	.3**	.17
Advanced Sales Training	.6**	.32*

* $p \leq .05$

** $p \leq .01$

13. Finley, R. Jr. (1970). Evaluation of behavior predictions from projective tests given in a management assessment center. Proceedings, 78th Annual Convention, APA, 567-568.

Type of assessment: Standard Oil of Ohio, 5 day assessment center; 5 man assessment committee (3 managers, 2 psychologists), groups of 12 assessees, 2 days of assessment

Tests: 1) Background information, 2) projective (TAT and incomplete sentences), 3) manager's interview, 4) formal talk by assessee, 5) 3 simulated business exercises, 6) intellectual abilities tests, 7) sociometric questionnaire

Predictive validity: Comparing assessment prediction ratings from projective reports for 13 psychological variables with 1) the assessment committee's predictive ratings based on all information sources (composite prediction) and 2) with the supervisor's on-the-job follow-up ratings several years later.

Two sample groups: Old (1): N=1109
New (2): N=119

Correlations of Projective vs. Supervisors' and vs. Committee Ratings and of
FACT Committee vs. Supervisors' Rating of Old⁽¹⁾ and Recent⁽²⁾ Samples

Rating Variable	Projective vs. Supervisor		Projective vs. Committee		Committee vs. Supervisor	
	(1)	(2)	(1)	(2)	(1)	(2)
Amount of Participation	.21*	.34*	.37*	.43*	.30*	.62*
Oral Communication	.07	.13	-.02	.35*	.23*	.39*
Personal Acceptability	.16	.32*	.38*	.45*	.29*	.31*
Impact	.15	.34*	.27*	.45*	.20*	.36*
Quality of Participation	.13	.25*	.22*	.31*	.22*	.25*
Personal Breadth	.30*	.12	.18*	.48*	.28*	.42*
Orientation to Detail	.09	.01	.25*	.60*	.00	.19*
Self-direction	.11	.07	.39*	.55*	.20*	.26*
Relation to Authority	.11	.21*	.70*	.62*	.15	.31*
Originality	.28*	.29*	.31*	.61*	.29*	.43*
Understanding of People	.11	.16	.39*	.50*	.17	.40*
Business Drive	.10	.02	.53*	.65*	.26*	.21*
Potential	--	.39*	--	.54*	.65*	.63*

*p = .05, 1T

Correlations tended to be high with the assessment committees ratings, but low with later supervisor's judgments. Committee ratings also correlated highly with supervisor's rating. Better results for more recent samples were attributed to refined procedures as well as shorter time interval to criterion.

APPENDIX C

In-Depth Look at one Management Assessment Program

It is useful to view a single, typical, assessment program in considerable depth (Townsend and Alderson, 1985). This examination will provide the reader with an detailed look at one assessment program, what its elements are, why they are there, and how they are related to each other in the assessment and feedback process.

This assessment program is conducted by field assessment centers operating under the general supervision of an international headquarters assessment and administration unit. Each field center is the office of a licensed psychologist who conducts standard assessments after being contacted by the personnel office of companies from the international conglomerate. Discussion of this program will include detail on the following:

- Participants
- Purpose of Assessment
- Procedures
- Tests Used
- Position Descriptions
- Assessors
- Costs

Participants

Professionals and executives from 32 requesting companies completed executive level assessment. The ratio of males to females was 12 to 1 with a mean age at the time of assessment of 37.2 years. There were slightly fewer applicants for hiring selection than there were internal candidates for promotion, transfer, or general evaluation. Percentages are used in the following tables because of the unequal N's contributing to the data. The 32 companies fell into the following nine general categories that reflect their business orientation or function:

<u>Company Type</u>	<u>%</u>
Electrical/Electronic Engineering	17.1
Insurance	16.2
Food Processing	15.9
Hotel	15.9
Mechanical Engineering	11.1
Publishing	10.2
Computer Software	6.0
Construction/Real Estate Development	5.7
Other (Headquarters, etc.)	1.8

Five areas of professional career experience were identified:

<u>Career Area</u>	<u>%</u>
General Business Management	39.9
Engineering/Technical	20.1
Marketing/Sales	19.8
Accounting/Finance	13.2
Legal	6.9

Participants were assessed for positions from lower level management to executive level management as follows:

<u>Level of Management</u>	<u>Grades</u>	<u>%</u>
Lower Middle	10-14	12.9
Upper Middle	15-18	57.7
Executive	19-24	29.4

Participants had work histories of from 2 to 39 years with a mean workspan of 14.5 years. Participants recorded having held between 1 and 16 previous positions with a mean of 5.5 jobs. The average number of companies previously worked for was 3.2.

Their educational experience represented a range from no college to Ph.D.s, with a mean educational level of 16.4 years. Sixty-three had Master's degrees and eight had Ph.Ds. The following table summarizes subjects by the highest level degree which has been received and by the declared major field of study:

<u>Degree</u>	<u>%</u>
Ph.D.	2.4
Law (J.D.)	8.7
M.A., M.B.A.	18.9
B.A., B.S.	50.2
None	19.8

<u>Field</u>	<u>%</u>
Business	33.1
Law, Economics, Political Science	17.0
Engineering/Technical	15.5
Accounting/Finance	10.3
Liberal Arts	5.8
Science/Mathematics	4.0
Life Sciences	3.6
None or not identified	10.6

Purpose of Assessments

Assessment is one of the three criteria for selection. The individual's work record provides the most important input. The company interview conducted by internal managers is the second critical basis for arriving at a decision. It is only after an employee or candidate has passed the first two criteria that an assessment meeting is arranged.

The decision to hire, promote, or transfer an individual, then, is made on the combined judgment of the record of the person's job performance, company interview, and an impartial assessment of the candidate from a psychological point of view. If a candidate is selected and the assessment recommendation has been favorable, it is because it concurs with the judgments based on the first two criteria. Likewise, if a candidate is rejected and the assessment recommendation is unfavorable, it is most likely that performance for one or both of the other criteria was questionable or marginal. Assessment judgments can be and sometimes are overruled by the company. Sometimes, in retrospect, the company's judgment was correct, and often, we have been led to believe, the assessment judgment was correct. Most of the candidates are in middle management, and some are at the entry level into executive ranks. None can join the executive ranks without an assessment evaluation.

One function of the assessment is to evaluate the individual in terms of the likelihood of success in the targeted position. In this case, the individual is assessed against position requirements and the norms of a comparable career group.

Another important function is to evaluate the candidate in terms of potential for rising to the executive ranks. It is for this reason that all candidates are compared to the executive level norms. In this context, the candidate is evaluated in terms of how nearly he/she possesses the psychological resources to meet the demands of an executive level position. It is in this context, also, that major assets and limitations may be described.

A third function, albeit a lesser one, of this assessment program is to identify limitations for the next level position or potential limitations at higher levels which may be amenable to training or developmental efforts. However, the assessment program as it is currently set up is not particularly suited to identifying developmental needs.

A final function of the assessment is to provide insight into the individual's strengths and weaknesses. The individual is evaluated for the cognitive or personality resources that are most likely to be drawn upon in meeting the demands of a position.

Procedure

The program spanned a ten-year period, with candidates each participating in a one-day executive assessment. Assessments were given to individuals being considered for employment (external candidates) or for advancement or transfer (internal candidates). Formal psychological assessment is a requirement for entry into the executive ranks of the company. Each candidate

was given the same battery of tests, administered by the same assessor who was assisted by a research assistant/ intern. The assessments were conducted according to professional standards and procedures. An eight-hour day was spent by each candidate taking eight pencil-and-paper tests:

1. Guilford-Zimmerman Test of Verbal Comprehension (GZ1)
2. Guilford-Zimmerman Test of Quantitative General Reasoning (GZ2)
3. Watson-Glaser Test of Critical Thinking (WG)
4. Productive Thinking Test (PT)
5. Guilford-Zimmerman Temperament Survey (9 scales)
6. Sentence Completion Test (SC)
7. Thematic Apperception Test (TAT) (cards #1, 2, 7BM, 12M, 17BM, 20)
8. Letter-writing exercise

In addition to the testing, each candidate provided a resume including work experience and educational background. The tests measured elements of intelligence and personality which are considered relevant to the requirements or demands of the position for which the individual was being considered. The requirements are defined by the position descriptions provided by the company and by informal discussions with the personnel directors.

Candidates are scheduled for assessment by a call from the company personnel office to the assessor. At the time of the appointment, the name of the candidate, the job title under consideration, grade level, and whether the individual is an internal or external candidate is recorded by the assessment office. If it is an internal candidate, it is determined whether the assessment is for general potential for upward movement, or for a targeted position or positions. Occasionally, a candidate may be seen as likely for one of two positions, with the company interested in evaluation for the better psychological fit. The company is requested to send a position description and, particularly, to indicate if there are any special concerns about the candidate in relation to the job or job demands. The more specific the information the company can provide about the position or special considerations, the clearer and easier is the task of assessment. At this time, the company also indicates whether it requires an immediate feedback by telephone, in response to some time urgency for information, or the regular feedback by written report.

Assessment Activities

In this program, the participants are assessed individually with individual tests. There are no group exercises. The letter writing exercise can be considered a situational test. It describes a work situation with sensitive interpersonal relationships; the participant is asked to draft a letter representing how he/she would typically respond to such a situation.

In addition to the numerous paper and pencil tests, each participant has two interviews. The first is fairly short, about one-half-hour at the beginning of the assessment day, during which the individual is informed of the general purposes and procedures for the assessment and any questions are answered. There is no systematic observation of the candidate during the interview other than an overall first impression. The interview is

semi-structured in that the same basic information is covered for each candidate; the candidate may raise questions, but an extended, open-ended discussion is not encouraged.

After the briefing, the candidate is taken to a testing room, usually a small library or an empty classroom. After the first two tests, which are timed, the remaining six paper-and-pencil tests are given to the candidate. The instructions for each are given and a recommended time and order for completion is stated. The candidate is requested to bring each test to the assessor's office upon completing it so that scoring may begin immediately.

Candidates are encouraged to adhere to the time allotments, but this is not over-stressed. By permitting candidates to work at somewhat their own pace, additional information is gained about their normal workstyle. The assessors are available to the candidates throughout the day to answer questions, but there is no directed, systematic observation of the candidate's behavior.

Opportunities for informal behavioral observations are relied on heavily and tend to be quite revealing. By the end of a long day of testing, the candidate's prevailing characteristics are usually evident. When behavior observations collaborate objective testing, they help to validate the testing. When behavioral observations conflict with objective testing, several options are suggested: 1) the individual's self-perceptions as indicated by the objective test may not be accurate, 2) the individual may be role playing in a manner that is not consistent with his/her psychological underpinnings, 3) the individual may have attempted to give an untrue representation either in the tests or in behavior (i.e., how he/she thinks he/she should be rather than how he/she really is). It then becomes the job of the assessor to determine what evidence to believe. The general rule of thumb is that unless a behavior is evidenced from two sources, it should be discounted as being characteristic of the individual. That is, for observed behavior to be believed, it must be confirmed either in the objective tests or in the subject's own words from the subjective exercises. Or, if the behavior is not directly observed, it must be indicated by both the objective and the subjective tests.

When the final exercise has been completed, the candidate is asked to take a half-hour break while the last exercise is scored and the data are summarized. The second interview, the feedback session, takes about one hour at the end of the day. During this time, the assessee is given the results of the tests in general terms. Since the assessors have not had sufficient time to fully integrate the findings of the day's exercises, the terms used to describe results are phrased to indicate their direction without being too specific. The feedback interview is unstructured and offers an opportunity for the assessor to clarify tentative conclusions and to resolve conflicting information.

Test Scoring

The objective tests provide the most consistent and reliable measures related to the attributes in the position descriptions. The tests of verbal comprehension, quantitative general reasoning, and critical thinking provide a

basis for evaluating most of the intellectual requirements of positions. The productive thinking test addresses the remainder. The productive thinking test is not an objective test like the others, it does not have a structured key for scoring subject's responses, but is scored according to a consistent set of criteria. It is considered a reliable contribution to the evaluation of cognitive factors. The four intellectual tests together address the cognitive requirements of the position descriptions.

The temperament survey provides nine objectively scored measures of personality variables that are informative in dealing with work approach and relationships with others. The scales that are used are general activity level, restraint, ascendance, sociability, emotional stability, objectivity, friendliness, thoughtfulness, and personal relations.

The four objective tests are scored by standardized keys against related corporate norms. Norms also exist for middle managers and several occupational categories, but these are not used unless specifically requested or unless such a comparison would, so identified, be more favorable (i.e., in identifying where there might be a better job fit). Executive norms are the basic norms against which all candidates are held.

The other four tests (productive thinking, letter writing, sentence completion, and the TAT) are subjectively scored without standardized keys. When two assessors are involved in the testing, each submits independent subjective scores. Differences are either averaged or a conference is held to reconcile differences so that a single judgment is produced reflecting both opinions.

Reliability checks, conducted occasionally and informally, indicate a high correspondence between the scorers. Whenever reliabilities begin to slip, it is generally an indication that it is time to review the criteria for judgment and sharpen skills. Structured or semi-structured scoring sheets for the subjectively judged tests are of enormous help in evaluating these tests. At the present time, only one of the tests (letter writing) has such a scoring sheet. Developing others for the remaining tests has been frequently discussed but not yet done by this assessment team. Scoring forms may well have been developed in other assessment centers. It should be noted that such scoring aids ideally should be flexible enough to note additional information and should be updated from time to time as needed.

Norms

There are three norms against which the candidate's performance is evaluated: 1) the corporate norms representing over 8000 executives who are in positions at grade 19 (entry level executive ranks) and above (to a theoretical grade 45); 2) the individual's own norms, that is, their own relative strengths and weaknesses on similar or related dimensions; and 3) to subnorms, which are less clearly specified than corporate norms in many areas, for particular companies and/or middle management career areas, i.e., engineers in a high tech company.

Specifically, test results are compared to requirements in the position description of current and future positions. This comparison shows in which areas the candidate's ability exceeds requirements in the present position and

which areas are appropriate for remedial training. Comparison to future position requirements shows the extent of development necessary to expect successful performance after promotion. The quality of the position description that is provided by the company, plus any additional information from the personnel office or director expressing specific concerns about the individual or requirements, or characteristics of the job situation, has a great influence on the ability of the assessors to evaluate the information provided by the tests.

Company Feedback

Immediate feedback, when requested, is usually given within 24 to 48 hours after the assessment. An immediate report does not replace the normal feedback report, but provides the company with a general summary of the test results and the overall recommendation when the company is under some time pressure to make a final decision. While an immediate feedback report necessarily is less specific than the regular report, no report is made until the assessor has summarized the results sufficiently to indicate whether the evaluation is favorable or not and to define the major findings. The immediate feedback never is reversed by the final report. The final, written report, which is submitted to the company within five working days after the assessment, further integrates the findings and documents the evidence upon which the conclusions were based.

Written Report

The written report is sent to the requesting officer, usually the Director of Personnel. The test results, the psychological profile, and copies of the written report are sent to the corporate assessment center where the information is used for establishing or updating norms or for other confidential research purposes. The requesting company receives only a seven or eight page narrative report summarizing the assessment findings. The report is descriptive in nature and is comprised of the following parts: Work Approach, Intellectual Effectiveness, Relationships with Others, Primary Assets, Primary Limitations, Summary Integration and Recommendations, and a background resume.

Test results are reported to the company using only general descriptive categories such as above average, average, low, etc. Number values are not recorded in the written report as they are considered too easily misinterpreted by individuals who are not trained in testing and evaluation.

Assessment information is generally considered valid for about three years. Some shifts of scores may be evident over time. However, it is rare to find a profile that has undergone major reversals.

The written report is retained by the Personnel Office. The assessment reports are used for making career decisions by the responsible company officers. Access to the reports for other purposes is determined by company policy, but generally confidentiality rules are strict.

Position Descriptions

Assessments are usually conducted with a specific job in mind for the candidate. The more information about the job, its work demands, and situational characteristics, the better will be the evaluation. A well-structured position description plus situational information give useful information about the kinds of psychological demands that will be placed on the individual. The assessment, in turn, can evaluate to what extent the individual has the psychological resources to meet those demands. To be poor in resources is not to say that the individual cannot do the job, only that he/she will have to reach further, train harder, and practice more in order to exhibit and maintain the desired behaviors. A poor position description or none means that the assessors must consider every dimension equally regardless of its relevance to the demands of the job.

Over 160 different position descriptions provided by the companies have been reviewed and categorized as part of the on-going research. The position descriptions were grouped according to their professional functions, i.e., finance/accounting or sales/marketing. Each of the position descriptions of the professional career areas was sorted into management level from pre-management to executive level. Comparisons of work activities and requirements were made to assure that positions were equivalent across companies and professional areas and to note any discrepancies between managerial levels and the functional responsibilities assigned to them.

Each position description below entry level executive were sorted into one of the four professional categories: 1) general business management, 2) finance and accounting, 3) sales and marketing, and 4) engineering and technical management. The position descriptions for each category were analyzed to identify the functions, activities, and job demands that were common to all positions, then analyzed to identify the characteristics that were distinctive to particular positions. Key words in the position descriptions were used to identify or infer the required psychological components in terms of psychological profile variables.

Each category of position description was summarized to describe the relevant intellectual demands, work approach, and relationships with others, from the characteristics that could be inferred as most likely to result in the desired behavior to carry out the demands of the position.

An effort was made to relate the position descriptions to Levinson's personality characteristics of an executive. Generally speaking, the position descriptions do a very poor job of addressing many of the characteristics on Levinson's scale. Of 20 items listed by Levinson, the position descriptions generally addressed about half. The ones addressed were essentially the same as the cognitive and temperament dimensions from the assessment materials.

A similar procedure was carried out for the analysis of the upper-level management position descriptions. These positions were found to fall into two major categories which overlapped greatly: 1) positions of command over production, technical, or human resources, and 2) positions that combine command functions and technical knowledge based functions. There is a third,

smaller category of strictly advisory positions based on specific or technical knowledge. The executive level technical advisor without command functions is not representative of most executives and was considered to present a different executive type.

The analysis and summarization of position descriptions for four middle management categories yielded four theoretical descriptions exemplified by the technical/engineering managers.

Engineering/Technical Management

Intellectual Demands. An engineering or technical degree is required and a quantitative base of knowledge is presumed, although position descriptions do not particularly specify quantitatively-oriented tasks. The professional is expected to be able to speak technically with peers and subordinates, but also to be able to explain technical information to non-technical peers and superiors. Report writing is a part of some staff-oriented positions. The major demand for quantitative skills lies in keeping abreast of the technical field by noting changes in the field and the activities of competitors. Professionals are expected to be critical and analytical in their evaluation of production functions. There is some human relations problem solving, but most problem solving is technical.

Work Approach. Professionals are given broad functional latitude. Their functions tend to focus either on managing production operations or in managing other engineers, and sometimes both. Professionals may be involved in planning and coordinating across operations, although the major functions focus on a single subject area or product. When working with professional technical peers, a diffused team effort is suggested with much emphasis on coordination and liaison. In some cases, the professional may be expected to act as a change agent, i.e., bringing in a new technology. Keeping up with the field is expected. Liaison across operations provides a relatively broad view of the company. Ingenuity in solving problems, particularly technical ones, is encouraged.

Relationship with Others. The professional is expected to be sensitive to personnel and labor issues, especially if directly managing others. Peer management is expected to be team oriented. Management of subordinates is expected to be directive, maintaining technical standards. Human problem solving may be either on a daily or occasional basis. Communications are important; professionals are expected to be articulate and effective with non-technical types, even on technical subjects. There is much communication with vendors and marketing persons.

The analysis and summation of the executive level positions yielded the following description which is the basis for a theoretical executive profile.

Executive-level Manager

Intellectual Demands. Executives deal with complex work and issues involving analysis and review of mid- to long-range plans. Creativity and flexibility are encouraged in the identification of problems and their solutions. They are expected to display shrewdness and good business acumen and to communicate well and effectively.

Work Approach. Their major responsibility is for the success of the functions to which they have been assigned. Executives may have to handle controversial assignments and defend unpopular positions with poise. They are goal directed, showing initiative and visible leadership. Activities are directed and coordinated by executives with the tasks delegated to subordinates. Executive leaders are expected to motivate their subordinates, while also commanding their respect. A major concern is for the development and utilization of human resources. Executives serve as liaison across the functions for which they are responsible, as well as with corporate decision makers.

Relationships with Others. Executives are expected to be assertive, firm, objective, yet understanding in their relationships. They are expected to support their subordinate staff in maintaining standards, to maintain good relations with decision makers, and to remain diplomatic when dealing with controversial areas. Internal and external contacts are extensive.

Level of Assessor Skill Required

Until recently, the assessment center for the headquarters conglomerate was headed by a licensed (clinical-industrial) psychologist. By corporate policy, all assessments are conducted under the direct supervision of a licensed psychologist. Auxiliary assessors need not be licensed psychologists, but a period of training is required to adequately interpret tests, integrate materials, and prepare reports. The length of training varies with the intensity of the effort and the complexity of the acquired skills. However, it is not unreasonable to say that a minimum of six months to a year of supervised training is required for a novice to become proficient at the required skills. Less complex tasks of scoring tests and systematic behavior observation require less training time.

Costs of Assessment

The professional services fee for a one-day assessment including the feedback interview and the written report to the company is \$485.00. Other costs incurred by the company are for transportation and hotel accommodations if the candidate is from out of town. If the candidate is from within the company, there is the added indirect cost of time away from the job. If a second feedback to the candidate is required, there is an additional fee of \$85.00. Company policy determines whether additional transportation and accommodations expense are incurred.